Department of Political and Economic Studies
University of Helsinki
Finland

ENDEAVOUR TO FIND EVIDENCE

THE ROLE OF EVALUATION IN COMPLEX SYSTEMS OF GOVERNANCE

Petri Uusikylä

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Social Sciences
of the University of Helsinki, for public examination in the U35, AUD 116,
on the 15th of November 2019, at noon.

Helsinki 2019

Valtiotieteellisen tiedekunnan julkaisuja 130
Publications of the Faculty of Social Sciences 130

Distribution and Sales
shop.unigrafia.fi

ISSN 2343-273X
ISSN 2343-2748

ISBN 978-951-51-3415-8 (pdf)
ISBN 978-951-51-3416-5 (nid)

Unigrafia
Helsinki 2019

ABSTRACT

The nature and logic of societal decision-making have changed from the rational planning era of the 1970s and 80s, through the performance management of the 1990s and 2000s, toward a multilevel system of governance today. Global economic and political issues are highly interconnected and complex. The speed of decision-making and the urgency of the issues being dealt with have increased. This has led to a situation in which the boundaries between policymakers, civil servants and external interest organizations are becoming more blurred.

Policymaking is now expected to be more open, transparent and evidence-based or -informed than in the past. There is a struggle to guarantee professional, high-quality and evidence-based policy preparation in a complex operating environment that is constantly changing. A role seems to have emerged for evaluation in supporting evidence-based policies in complex settings.

This dissertation examines the changing role of evaluation in complex systems of governance. The problem with current evaluation approaches is that they are rather static; apply mechanistic, linear causal logic; lack understanding of the theory of change; and often rely on rigid, *a priori*, defined evaluation criteria and methods. Traditional evaluation models are not suitable for understanding and evaluating reasonably complicated or complex policy phenomena.

The dissertation outlines an alternative systemic evaluation approach that is more suitable for evaluating the effectiveness and impacts of complex polices and interventions than the traditional policy or program evaluation model. The new approach presented relies heavily on the idea of systems and complexity theories and the utilization of the systemic evaluation designs derived from them. Developmental evaluations, systemic network analysis and utilization of theory of change models are examples of this approach. The thesis also suggests that the evaluators should have a more active role as knowledge brokers and policy interpreters between governments and citizens. This would make it easier for citizens and firms to understand the complex regulatory and policymaking environment, and thus to support the development of an open and democratic society.

Keywords: evaluation, governance, complexity theory, systems thinking, causality

TIIVISTELMÄ

Yhteiskunnallisen päätöksenteon luonne on muuttunut 1970- ja 80-lukujen rationaalisesta suunnittelusta 1990- ja 2000-lukujen tulosohjauksen kautta kohti monitasoista ja systeemistä hallintakokonaisuutta. Globaalissa taloudessa ja politiikassa ilmiöiden välillä vallitsee vahva keskinäisriippuvuus ja yhteenkietoutuneisuus. Kun samaan aikaan päätöksenteon vauhti ja käsiteltävien asioiden kiireellisyys ovat kasvaneet, ollaan jouduttu tilanteeseen, jossa päätöksentekoa ei enää ohjaa selkeä työnjako poliittisten linjausten sek**ä hallin**nollisen valmistelun ja toimeenpanon välillä.

Samaan aikaan päätöksenteolta odotetaan yhä suurempaa avoimuutta ja läpinäkyvyyttä sekä vahvaa tiedolla johtamista. Ongelmaksi on noussut huolellisen ja näyttöön perustuvan valmistelun yhteensovittaminen jatkuvassa muutoksessa olevaan monimutkaiseen ja kompleksiseen toimintaympäristöön. Mikä on arvioinnin rooli ja mahdollisuudet näyttöön perustuvan politiikan tukemisessa?

Väitöskirjassa tarkastellaan arvioinnin muuttuvaa roolia kompleksisessa hallintajärjestelmässä. Nykymuotoisen arviointitoiminnan ongelmana on liiallinen staattisuus, mekanistinen lineaarinen kausaalilogiikka, puutteellinen muutoksen teorian ymmärrys sekä jäykät usein *a priori* **määritellyt arviointikriteerit ja niihin liittyvät menetelmät. Perinteiset arviointikäytännöt soveltuvat kohtuullisen huonosti monimutkaisten tai kompleksisten politiikkailmiöiden ymmärtämiseen ja arviointiin.**

Tässä väitöskirjassa hahmotetaan vaihtoehtoista, systeemistä arvioinnin lähestymistapaa, joka soveltuu perinteistä politiikka- tai ohjelma-arvioinnin mallia paremmin kompleksisten hallintajärjestelmien vaikuttavuuden arviointiin. Esitetty uusi lähestymistapa nojaa vahvasti systeemi- ja kompleksisuusteorioiden ajatukseen ja niistä johdettujen systeemisten arviointiasetelmien hyödyntämiseen. Esimerkkeinä tällaisesta voidaan mainita kehittävä arviointi, systeeminen verkostoanalyysi, muutoksenteorioiden hyödyntäminen jne.

V**äitöskirja esittää** myös, että evaluaattoreilla tulisi olla nykyistä aktiivisempi tiedonvälittäjän ja politiikkatulkin rooli päätöksentekijöiden ja kansalaisten välillä. Tämä olisi omiaan lisäämään kansalaisten ja yritysten ymmärrystä monimutkaisesta sääntely- ja päätöksentekoympäristöstä sekä tätä kautta tukemaan demokraattisen järjestelmän avointa ja kansanvaltaista kehittämistä.

Avainsanat: Arviointi, hallinta, kompleksisuusteoria, systeeminen ajattelu, kausaliteetti.

ACKNOWLEDGEMETS

It has been an extraordinary privilege to work on research that I find both interesting and consequential. Undertaking this PhD has been a truly life-changing experience for me. However, it would not have been possible to complete without the support and guidance that I received from my mentors, colleagues, co-workers and other evaluation professionals.

First, I would like to very warmly thank my thesis advisors, Prof. Pertti Ahonen and Prof. Mikko Mattila, for their support and encouragement during the many months I spent finalizing my dissertation. Without guidance, motivation and constant feedback from you Pertti, this PhD would not have been achievable. I am also extremely grateful to my preliminary examiners, Prof. Antti Syväjärvi (University of Lapland) and Prof. Jari Stenvall (University of Tampere), for their careful reading of and constructive feedback on the draft version of my thesis.

When I started my professional career and research in the field of evaluation more than 25 years ago, I did not realize the extent to which research involves teamwork. Petri Virtanen (PhD) has been a great friend, colleague (at both the Ministry of Finance and Net Effect Ltd.), co-author and business partner. He has motivated me to finalize this dissertation, served as a good example by publishing a vast number of high-level academic books and articles, given me inspiring new ideas, helped me refine existing ones, and asked questions I would have never thought to consider. Ville Valovirta, my close friend, fellow member of a book club, and co-author of one of my articles, has also been an incredibly important colleague whose ideas and feedback have shaped this research from the beginning.

My colleagues in the Political Science department at the University of Helsinki have been essential to the completion of this research, as they made available resources that I did not know existed and solved problems that I could not have managed on my own. Also, thanks to Prof. Hanna Wass and Prof. Anne Holli, whose guidance was extremely helpful as I shaped the dissertation to fit the faculty's requirements. Furthermore, I am thankful for the helpful feedback and comments provided by the other PhD students in the Political, Societal and Regional Change doctoral program.

Although I was not yet aware, my interest in systems thinking started in early 1990s when I started to explore the science of social networks with two of my colleagues, Jan-Erik Johanson and Mikko Mattila, at the University of Helsinki. That was an unforgettable journey, and so was our trip to the INSNA Sunbelt conference to New Orleans in 1994. We were really ahead of our time; even at the dawn of the Internet and social media platforms, we intuitively felt that social

networks would shape political, organizational and social life in the future. Thank you Jani and Mikko.

I greatly appreciate the support and innovative ideas I received through collaborative work undertaken with the Finnish Complexity Network and Complexity Group at the University of Vaasa, especially with Harri Raisio and Pirkko Vartiainen. Additionally, my discussions with Harri Jalonen, Mikko Kosonen, Kirsi Hyytinen and Petri Salo have been inspiring and intellectually valuable, and they have shown me new horizons regarding systems change, governance and complexity thinking.

I am thankful for the support and learning opportunities I received from the collaborative work I undertook with my colleagues at Net Effect, the Finnish Evaluation Society and the Finnish section of the IIAS. I am especially grateful to Maria Suoheimo at the Finnish Red Cross, Katri Vataja at Sitra, and Venla Voutilainen at the Ministry for Foreign Affairs for giving me the opportunity to test systems evaluation methods in our joint evaluation projects. Network analyses during my field trips to the Philippines and Vietnam would not have been possible without the collaboration of my great colleagues, Paula Tommila and Tapio Wallenius. Also, very special thanks to Johanna Jokinen and Alistair Heather for their translations and proofreading.

I would like to give heartfelt thanks to my mother, Anna-Liisa, and father, Antti, for always believing in me and encouraging me to follow my dreams. I am also grateful to my brother, Timo, who helped me to understand the dependent source of interconnectedness and how "the Mystery of all mysteries is the door of all essence". Further, I am thankful to my wife, Anna, who has lived in the joint field by my side throughout my time in the PhD program and shared my life path. Without her, I would not have had the courage to embark on this journey in the first place. Last, but certainly not least, loving and warm thanks and hugs to my two gorgeous and brilliant daughters, Ida and Rosa, with whom I have had so many inspirational discussions and who gave me a final reason to complete this work.

LIST OF FIGURES AND TABLES

LIST OF FIGURES

Figure 1. Composition of the dissertation.

Figure 2. The transformation from a compliance-based system to a performance-oriented system

Figure 3. VUCA-dimensions and their main elements.

Figure 4. Evaluation linked with the policy cycle.

Figure 5. Macro-proposition: public management reform increases efficiency or innovativeness.

Figure 6. Macro- and micro-level propositions: effects of public administration reform on a public management system.

Figure 7. Quasi-experimental evaluation design

Figure 8. A linear evaluation model applied by the European Commission.

Figure 9. A causal influence chart.

Figure 10. The true and false social intervention logic of a program.

Figure 11. The intervention logic of the Logical Framework Approach and Systems Approach.

Figure 12. Developmental evaluation as dynamic reframing.

Figure 13. When to choose developmental evaluation approach.

Figure 14. Caloocan multi-risk map of interconnectedness.

Figure 15. Inter-organizational DP/DRR Network in Caloocan City

Figure 16. Collective Actors´ Roles in the Philippines DM Network

Figure 17. A Synthesis: Critical elements for coping with complexity - TAE.

Figure 18. Multi-level perspective on complex governance.

Figure 19. Mission-oriented policies and adaptive programming model

Figure 20. Trade-off in adaptive programming: Accountability vs. Innovativeness

Figure 21. The nature of problems or program design vs. evaluation logic.

Figure 22. Evaluators taking part in strategic alliances.

LIST OF TABLES

Table 1. Four Models of Governance

Table 2. Cynefin Framework and alternative evaluation designs

Table 3. Differences between traditional Program Approach and Systems Approach

Table 5. Main evaluation areas of complex adaptive policy systems

Table 6. Aklan Province Outcome Harvesting Results

ABBREVIATIONS

BSC	Balanced Scorecard
CA	Contribution Analysis
CG	Collaborative Governance
CMO	Context, mechanism, outcome
DAC	Development Assistance Committee
DE	Developmental Evaluation
DM	Disaster Management
EBP	Evidence-based Policymaking
EC	European Commission
E&I	Entrepreneurship & Innovation
ESG	Environmental, social and corporate governance
EU	European Union
FRC	The Finnish Red Cross
GII	Global innovation index
GDP	Gross domestic product
GoF	Government of Finland
DP	Disaster Preparedness
DRR	Disaster Risk Reduction
HEI	Higher Education Institution
HQ	Head Quarter
IFRC	International Federation of Red Cross Red Crescent
IG	Interactive Governance
KPI	Key Performance Indicator
KTS	Keskipitkän aikavälin suunnitelma
LFA	Logical Framework Approach
LGU	Local Government Unit
MFA	Ministry for Foreign Affairs
M&E	Monitoring & Evaluation
NAO	Network Administrative Organisation
NG	Network Governance
NGO	Non-Governmental Organisation
NIS	National Innovation System
NPM	New Public Management
NWS	Neo-Weberian State
OECD	Organisation for Economic Co-operation and Development
OH	Outcome Harvesting

OM	Outcome Mapping
PA	The Participatory Approach
PAA	The Policy Advice Approach
PDIA	Problem-driven and Iterative Adaptation
PMEL	Planning, monitoring, evaluation, and learning
PPBS	Planning, Programming and Budgeting System
PRC	The Philippines Red Cross
PTS	Pitkän aikavälin suunnitelma
R&D	Research and Development
RBM	Results Based Management
RCT	Randomized controlled trial
RE	Realist Evaluation
RIM	The Risks Interconnections Map
RPA	Rational Planning Approach
SA	Systems Approach
SDG	Sustainable Development Goal
SNA	Social Network Analysis
S&T	Science and Technology
ST&E	Science, Technology and Engineering
SMEs	Small and Medium-sized Enterprises
TM	Transition management
ToC	Theory of Change
ToR	Terms of Reference
TPA	Traditional Public Administration
TPT	Traditional Program Theory
VCA	Vulnerability and Capacity Assessment
VUCA	Volatility, uncertainty, complexity, ambiguity
WEF	The World Economic Forum

CONTENTS

ABSTRACT 3
TIIVISTELMÄ 4
ACKNOWLEDGEMETS 5
LIST OF FIGURES AND TABLES 7
ABBREVIATIONS 8
1 INTRODUCTION 13
2 THE TRANSFORMATION OF SYSTEMS OF GOVERNANCE 17
3 EVALUATION AND EVIDENCE-BASED POLICY 26
3.1 Why Evidence-based Policymaking? 28
3.2 Public Management Reform as Evidence-based Policy 29
3.3 How Evaluations Support Evidence-based Policymaking? 35
4 SYSTEMS APPROACH AND COMPLEXITY 51
4.1 Systems Thinking and Emergence 51
4.2 Complexity Theory and Challenges to Evaluation 57
4.3 Evaluating Networks and Complex Adaptive Systems 61
4.4 Case Study: Systems Evaluation on Disaster Preparedness 67
5 SYNTHESIS: ADAPTIVE EVALUATION FOR TRANSFORMATIVE GOVERNANCE 78
5.1 Transformative Governance 79
5.2 Mission-oriented Policies and Adaptive Programming 81
5.3 Evaluators as Knowledge Brokers 86
5.4 Conclusions 92
REFERENCES 97

"The world we have created today as a result of our thinking thus far has problems which cannot be solved by thinking the way we thought when we created them."

- Albert Einstein

1 INTRODUCTION

The demand for evidence-based decision-making usually arises in the context of failed policy decisions. A recent example in Finland is the inability of the government of Prime Minister Juha Sipilä to finalize a social and health care reform that was one of the top priorities of the government agenda. There has been criticism from both the public and researchers that the government should have heard the views of experts, and based its decisions on reliable research findings. The requirement is certainly justified, but there are a number of ontological and epistemological problems with evidence-based decision-making that are often overlooked in public debate. These problems relate typically to demand and supply of evidence, cognitive limits and bounded rationality of policymakers and complexity and unpredictability of policymaking environment (cf. Cairney 2016, 13).

This dissertation provides a broader understanding of the role of evidence and evaluation in an effort to improve the knowledge base of policymaking and thus produce better policy outcomes. The roots of evidence-based decision-making in natural sciences, where the relationships between cause and effect are generally easily verifiable (e.g. smoking is hazardous to your health, the use of fossil fuels accelerates climate change, and eating fatty foods increases cholesterol and may block your blood vessels). To obtain evidence, clinical trials and measurements can be performed, and experimental settings can be designed in which the experimental group is exposed to an effect (for example, a new drug) and a control group is given placebo. Randomized controlled trials (RCTs) can provide reliable comparative information on the net effects of the intervention in question. Similar thinking has also been applied in social science research and public policy evaluations. Since the standardization of test and control groups is difficult and may lead to ethical research problems, social science scholars often use quasi-experimental designs in which the changes in the control group can be derived from statistical evidence.

This thesis challenges the above mentioned simple models of knowledge management and evidence-based decision-making. The main argument is that in order to create an evidence-base for decision-making and the procedures that follow, we need to understand the complexity, interconnectedness and nonlinear mechanisms that blur the chain of cause and effect. In a societal context, it is more suitable to use term evidence-informed decision-making rather than evidence-based decision-making (Bowen & Zwi 2005). Evidence-informed policymaking also accounts for other factors influencing the decision-making process, such as political expediency, values, feasibility, political agreements, bargaining and horse-trading, the influence of media and public debate, and other factors affecting the

decision-making situation on a case-by-case basis. In decision-making situations, it is also often crucial to have previous experience as well as tacit knowledge. Polanyi (1967) describes tacit knowledge as something that is acquired, in part, by practice and can be only partially communicated and shared. It often plays a significant role in decision-making situations.

This doctoral dissertation consists of four articles published in international journals in addition to this synthesis chapter. Each article supplements the argumentation related to the problems of mechanistic, linear causal reasoning in evaluation. In addition to the criticism of naïve causal thinking, the articles point out that if policies are evaluated through a single program or project perspective, the dynamics of the societal change will be missed and the overall understanding of the sustainable impacts of the policies cannot be achieved. Similarly, the impact of a single program on the phenomenon is difficult to evaluate. Thus, the effectiveness of the program is mixed with other intervening factors, which does not help program evaluation. Figure 1 presents the composition of this dissertation.

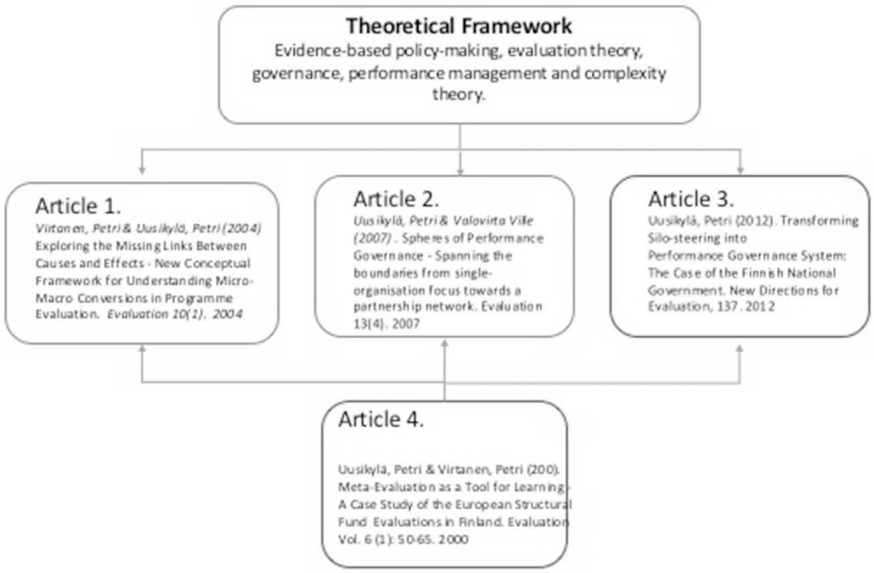

Figure 1. Composition of the dissertation.

Article 1. Virtanen, P. & Uusikylä, P. (2004). Exploring the Missing Links Between Causes and Effects - New Conceptual Framework for Understanding Micro-Macro Conversions in Program Evaluation. 77–91. *Evaluation* Vol. 10(1).

The theoretical and conceptual models included in the article were developed jointly by Petri Virtanen and Petri Uusikylä. Uusikylä and Virtanen wrote the article together. Petri Virtanen was the author responsible for the following

sections: introduction, causality assumptions and program theories. Uusikylä was responsible for the sections on the use of meta-theories to find causal explanations in the implementation of programs and the conceptual model for a research agenda. The authors wrote the chapter on evaluation paradigms and programs together. The authors revised the article jointly based on the comments of anonymous reviewers for the journal "Evaluation." Article discusses the ontological premises, conceptions of causality and relationships to the rational theories of action of different program evaluation paradigms.

Article 2. Uusikylä, P. & Valovirta, V. (2007). Three Spheres of Performance Governance - Spanning the Boundaries from Single-Organization Focus Toward a Partnership Network. 399–419. *Evaluation* Vol. 13(4).

The idea behind this article was developed jointly by Petri Uusikylä and Ville Valovirta. Uusikylä and Valovirta wrote the article together. Petri Uusikylä was the author responsible for the following sections: introduction, common problems in performance management, from performance management to performance governance, three spheres of performance and integrated framework for policy design and performance measurement. Valovirta was responsible for the sections of logic models and theory-based evaluations, contribution analysis, and balanced scorecard and cognitive mapping. The chapter on bringing logic and agents together and the conclusions were written together. The authors revised the article jointly based on the comments of anonymous reviewers for the journal "Evaluation."

Article 3. Uusikylä, P. (2013). Transforming Silo-steering into Performance Governance System: The Case of the Finnish National Government. 33–43. *New Directions for Evaluation*, 137. Wiley.

Petri Uusikylä was the only author of the article. He was responsible for developing the theoretical framework, empirical analysis and discussion. The article was reviewed by Steffen Bohni-Nielsen, David Hunter and two anonymous referees. The article discusses efforts by the Finnish central government to reform its approach to governing and service delivery through the implementation of performance management in various forms. In this context, the implementation and utilization of performance evaluation have been fragmented and narrow, resulting in the limited use of evaluative data in policy development and service improvement.

Article 4. Uusikylä, P. & Virtanen, P. (2000). Meta-Evaluation as a Tool for Learning - A Case Study of the European Structural Fund Evaluations in Finland. 50–65. *Evaluation* Vol. 6 (1).

This article was written by Petri Uusikylä and Petri Virtanen. Petri Uusikylä was the author responsible for the article. He was responsible for the meta-evaluation criteria and the empirical analysis. Petri Virtanen was responsible for the theoretical

foundations of the meta-evaluation. The authors wrote the conclusions together. This article aims at elaborating the approach of aggregating knowledge by introducing a more learning-oriented interpretation of the concept of meta-evaluation. In this sense, meta-evaluation should be part of an open dialog between various parties in the evaluation process.The summary part of the dissertation proceeds as follows. The second chapter discusses the change in public policy and governance from the point of view of societal steering, guidance and knowledge-based decision-making. The key observation is that, as we move from top-down steering to decentralized performance management and governance structures, the knowledge base behind decision-making also diversifies and becomes more polyphonic. This creates discontinuities in the preparation of the policy process, whereby the unity of policy content can be broken.

The third chapter examines the assumptions behind evidence-based or evidence-informed decision-making and the role of evaluation in the production of knowledge. The chapter highlights the constraints of utilizing research and evaluation data. The fourth chapter outlines an alternative evaluation perspective based on systems thinking and complexity theory. The chapter summarizes the possibilities and constraints of a complex evaluation perspective. The last chapter draws together the previous remarks and conclusions and outlines the new role of the evaluator as an information and knowledge broker. Finally, the topics for further research will be outlined.

2 THE TRANSFORMATION OF SYSTEMS OF GOVERNANCE

The purpose of this chapter is to describe the change in the management governance thinking behind evidence-based decision-making ideas. The knowledge base of the traditional planning paradigm was largely an *ex ante* assessment of future short, medium and long term needs and resources. The analysis was based on statistics, and this information coded into central and regional governments' centralized planning and budgeting systems such as PPBS or KTS/PTS in Finland (see Faludi 1978; Yewlett 1985; Pitkänen 1975; Tiihonen & Tiihonen 1990). Transition to performance management and decentralized management thinking also meant breaking down the centralized data bases. These were replaced by commonly defined result and impact goals, and monitoring and evaluation systems required for verification of the goal achievement in decentralized management were transferred to result-driven agencies and regional administrations.

Evidence-based policymaking (Davis et al. 2000; Segone 2004; Cairney 2016) lies at the heart of New Public Management (NPM) ideology. During the 1990s and early 2000s, most of the OECD countries carried out major budget reforms in order to stimulate and enhance their economies, efficiency and productivity, effectiveness and service quality. New Public Management is a results-based management framework for steering public service organizations and is used in government and public service institutions and agencies at both sub-national and national levels. According to Temmes (1998), NPM is a concept coined by researchers of administration, and its scientific basis rests partly on neoliberalism since it emphasizes the inefficiency of the welfare state and points out the red tape factors causing governmental failure.

However, many other scholars see NPM as a less politically laden concept of governance (Metcalfe & Richard 1990; Pollit 1993; Pollit & Bouckaert 2011). Evaluation forms a knowledge platform for evidence-based policymaking and is a critical precondition when transforming hierarchical planning systems into flexible and decentralized performance-oriented management systems. In performance management systems, detailed budget control and regulation are replaced by information steering and output control rather than by implementing strict control on inputs and usage. This requires not only a fundamental change of existing planning and steering systems but also a new management culture that focuses on results and impacts and rewards actions according to those. Figure 2. shows the main differences between the compliance-based system system and performance-oriented system. In the latter one the top-down command chain has been replaced

by the contractual arrangement. Also the regulatory emphasis of the compliance system is transformed towards information steering and other non-binding "soft" policy instruments.

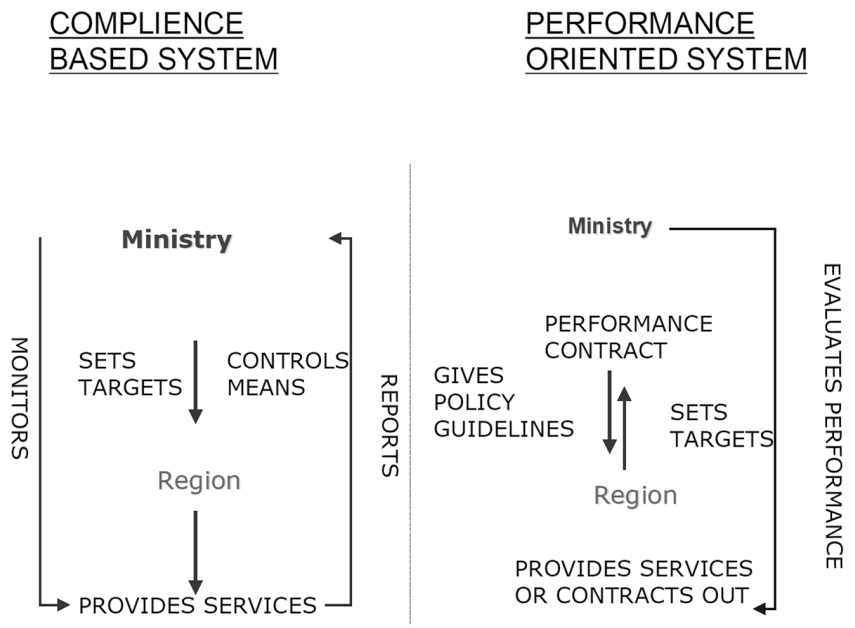

Figure 2. The transformation from a compliance-based system to a performance-oriented system (source: the author).

Evaluation should also be embedded within the policy cycle. In an ideal case, evaluation gives support to agenda setting (by highlighting societal problems), policy formulation and decision-making. These are typical tasks of *ex ante* evaluations or impact assessments. While policies are being implemented, evaluation and monitoring give accurate information to policymakers and program managers on how policies work (and, if they do not work, why). This support can be given by ongoing evaluation (either external or self-evaluations). Evaluation of results and impacts of the policies and program can only be verified some years after implementation. In the field of program evaluations, *ex post* evaluations are normally carried out two to five years after programs are completed. In policy evaluation, *ex post* evaluations are normally linked with the election cycle, i.e., the final evaluation of government policies is normally carried out at the end of the government period. This, however, is not likely to enhance policy learning and our understanding of institutional change and transformation.

Institutional change or reform policies cannot be explained by narrow, actor-oriented, rational choice models simply because they cannot give a satisfactory answer to the fundamental question: how do the interests of policy actors emerge,

and why do these interests change over time? Nor can one scrutinize the hollow institutional structures as natural building blocks of the welfare state, since institutions should always be seen as social outcomes that were created by a more or less rational decision process (Knight 1994). As Heiskanen and Martikainen (1988, 358) put it, "in such research, which aims at political and administrative explanation. The holistic perspective means that research results should not only reveal the relative importance of political and administrative causal factors but also provide theoretical insights into and interpretations of decision-making and implementation processes and their transformations in time."

As Ahonen & Salminen (1997) have demonstrated in their ambitious work "Metamorphosis of the Administrative Welfare State," the evaluation of institutional systems' change or reform is a complex and extremely difficult task. Administrative reforms cannot simply be assessed as "technical reforms" or new management models or tools introduced; reform policy is always embedded in the political-cultural context and guided through political rationality that does not always coincide with administrative rationality.

Institutional factors play a fundamental role in this analysis. On the one hand, the organization of policymaking affects the degree of power that any set of actors has over policy outcomes; on the other hand, the organizational position also influences an actor's definition of his own interests by establishing his institutional responsibilities and relationship to other actors (Hall 1986). It is, however, crucial to emphasize that institutions may constrain and refract politics but are never the sole cause of policy outcomes.

Also following Heclo's (1978, 306) concept of policy-oriented learning, policy change often refers to relatively enduring alterations of thought or behavioral intentions that result from experience and are concerned with the attainment of policy objectives. Such learning comprises, however, only one of the forces affecting policy change over time. In addition to this cognitive activity, there is a real world that changes as well. Policy networks also change traditional (or corporatist) policymaking culture and bring new players into policy domains such as social welfare and health care arenas (Mattila 2000).

Sabatier's advocacy coalition framework offers an interesting tool for combining changes in the belief system with external system events such as changes in the economic environment or in the governing coalition. The model of the advocacy coalition also fits perfectly with the network models used as the corner stones of our political constructivist model in this study. Sabatier (1993, 16) argues that "the most useful way to think about policy change over such a time span is through a focus on 'policy subsystems,' that is, the interaction of actors from different institutions who follow and seek to influence governmental decisions in a policy area."

Institutional change or transformation should not be treated simply as a managerial reform in a narrow, technical sense or something inevitable caused by technological innovations and free trade developments under the globalization

process. The role of evaluation should be reconsidered given changes taking place in twenty-first century democracies. The power shift itself raises new problems of governing modern societies. In this context, governance refers to all "the patterns that emerge from governing activities, (i.e. efforts to guide, steer and manage societies) of social, political and administrative actors. [...] Governability of a social-political system can be seen in terms of balancing process. It is not something static, but a constant process of coming to grips with the tension between governing need on the one hand (problem situations or the grasps of opportunities) and governing capacities (creating patterns of solutions or developing strategies) on the other hand." (Kooiman 1993, 2). Some see this emerging complexity merely as an inevitable trend of development (and, in fact, a desired one) that the state no longer governs through hierarchies, but rather steers actions in a more invisible way from a distance or by being the collaborator behind societal actors (Dunsire 1993; Tiihonen 2004; Klijn 2008). Some even yearn for the past and claim that the state as an authoritative governing body should be brought "back in" (Skocpol et al. 1985). It is obvious that the ungovernability of the system may cause some profound changes in public accountability and societal power structures and affect the legitimacy of the political system as a whole (Mayntz 1993, 9).

The increasing societal complexity and interconnectedness set new challenges both for the strategic management and evaluation of public-sector performance (see Virtanen & Uusikylä 2004). New strategic challenges have to do with managing complex inter-organizational networks, coping with complex and interconnected policies and creating proactive strategies rather than just reacting to environmental changes. As Kickert and Koppenjan (1997, 39) put it, "In modern society, an approach to public management not only has to deal with norms and values that go far beyond the criteria of effectiveness and efficiency [...] Public management is the 'governance' of complex networks, consisting of many actors [both private and public] [...] Public 'governance' is the directed influencing of societal processes in a network of many co-governing actors. These actors have different and sometimes conflicting objectives and interests." There is a clear need to move from performance management to performance governance.

Kooiman (1993, 2) defines governing as "all those activities of social-political and administrative actors that can be seen as purposeful efforts to guide, steer, control or manage (sectors or facets of societies" and governance as "the patterns that emerge from governing activities of social-political and administrative actors." Pierre and Peters (2000) treat governance both as a structure and process. They begin by discussing four common institutional models of governance: hierarchies, markets, networks and communities. Each of the four structural arrangements addresses the problem of providing direction to society and the economy in its own way. As Pierre and Peters (ibid., 15) note, each appears effective in solving some parts of the governance problem, but each also has its weaknesses. They also claim

that since each of the solutions is culturally and temporally bound, they may be effective in certain places and times, but are not a panacea for all societal problems. Virtanen and Stenvall (2018) discuss the role of policy intelligence in developing health services. The intelligent policy means that policymakers need to move towards a more comprehensive understanding about health. This is related to technology and digitalisation, which affect healthcare services. There is more and more co-creation of value, quality and efficiency. The third cornerstone, according to Virtanen and Stenvall (ibid.) is intelligent management and leadership of services. Table 1 summarizes alternative models of governance: 1) traditional public administration or Neo-Weberian State (the hierarchy model), 2) New Public Management (the market model) and 3) network governance (the network model).

Table 1. Four Models of Governance

	Traditional Public Administration (TPA)	**Neo-Weberian State (NWS)**	**New Public Management (NPM)**	**Network Governance (NG) and Collaborative Governance (CG) and Interactive Governance (IG)**
Context	Stable	Relatively stable	Competitive	Continuously changing
Core claim	To have a public sector–driven service system that provides services equally to all citizens	To modernize the traditional state apparatus so that it becomes more professional, more efficient and more responsive to citizens	To make government more efficient and consumer-responsive by market-type mechanisms	To make government better informed, more flexible and less exclusive by working through self-organizing networks
Needs/ problems	Straightforward, defined by professionals	Participatory, citizen-oriented	Wants expressed through markets	Complex, volatile and prone to risk
Improvement	Large step change improvement initially, less capacity for continuous improvement	Incremental changes that rely on strong public-sector institutions	Improvements in the managerial process and systems. Customer focus produces quality improvements in services	Aiming for transformational and continuous improvement in front line services
Population	Homogenous	Homogeneous	Atomized	Diverse
Role of policymakers	Commanders	Strong political steering	Announcers	Leaders and interpreters
Role of public managers	Executors and implementing agents	Implementing agents with strong professional capacity	Efficiency and market maximizers	Explorers
Role of population	Clients		Customers	Co-producers
Strategy	State and producer-centered	Citizen-oriented	Market and customer-centered	Shaped by civil society; co-production
Governance through actors	Hierarchies, public servants	Hierarchies with a responsive attitude toward citizens needs	Markets; purchasers and providers; clients and contractors	Networks and partnerships; civic leadership
Key concepts	Public good	Public good	Public choice	Public value
Some key sources	Weber	Dreschler and Kattel 2008; Lynn 2008	Hood 1991; Pollit 1990; Lane 2000	Kickert & Koppenjan 1997; Klijn & Koppenjan 2012; Provan & Kenis 2008; Ansell & Gash 2008; Torfing et al. 2013.

Source: Modified from Hartley 2005; Hyytinen 2017; Pollit & Bouckaert 2011.

Table 1 describes four governance models and the key factors that define them. The table is based on a summary by Hartley (2005), an application by Hyytinen (2017), and a classification by Pollit and Bouckaert (2011). Collaborative governance and interactive governance models have been added to complement network governance theories. The first model is a traditional public administration, i.e., Weberian bureaucracy. By bureaucracy, Weber meant a rational and objective leadership and management structure that is relatively stable. In a bureaucracy, the power of the leader is based on written rules and formal responsibility. The manager's personal status is separate from the position of the leader in the organization. It must be remembered that bureaucracy is an ideal type that rarely occurs in its pure form in real-life administration. However, bureaucracy is linked to the modern rational organization of work and activity. In this case, management is thought to be solved by objective rules. Compliance is the responsibility of qualified, specialized workers. Thanks to rules and the hierarchical organization, the organization can act rationally as an individual's preferences can be prevented. The neo-Weberian model (Dreschler and Kattel 2008; Lynn 2008) is an improved model of traditional public administration, i.e., Weberian model 2.0. Bureaucracy prescribes fixed relationships among positions through which incumbents flow, without affecting organizational operation. The idea of the neo-Weberian model is to maintain a strong state-orientation of traditional public administration and the primacy of public services. According to the model's supporters, through incremental reforms, a strong state model can be made more efficient, more professional and more responsive to the needs of citizens. The models of France, Germany and the Nordic countries were regarded as an example of the neo-Weberian model of governance until recently. However, budget cuts to public services, privatization and the use of different hybrid models (see Johanson & Vakkuri 2017) have challenged that argument.

The idea behind New Public Management is to make government more efficient and citizen or consumer responsive by applying bottom-up decision-making processes (Pollit 1990; Lane 2000). The idea of the NPM is to shift the focus away from input control to monitor and evaluate results and impacts and to use market-type mechanisms such as contracting out to enhance efficiency of public sector services. Model as such has enabled decentralized decision-making and been somewhat successful in many countries (especially in the Nordic countries). However, the implementation of the model has failed in many OECD countries due to problems of inadequate performance indicators, lack of reliable monitoring information and the weak guiding effect of the performance agreements. Criticism of the NPM paradigm began mounting in the early 2000s. The NPM was considered to be overly mechanical and top-down, along with placing undue emphasis on the performances of public management. At the same time, performance management had, despite its original goals, created an excessively differentiated and specialized

sector structure in governance, which was no longer capable of meeting the overall needs of an increasingly complex society or demonstrating genuine cost-effectiveness and impact. Therefore, it is no wonder that recent reforms have emphasized governance, networks, agility, phenomena, transformativity and the unity of the government as a whole (Christensen & Laegreid 2016; Virtanen et al. 2016).

The network model represents a new and agile way of defining the public sector's role and operating model. The idea is to make government better informed, more flexible and less exclusive by working through self-organizing networks. Provan and Kenis (2008) distinguish three basic models, or forms, of network governance developed according to their distinct structural properties. The first and simplest form of network governance is participant governance. According to Provan and Kenis (ibid.), "This form is governed by the network members themselves with no separate and unique governance entity. Governance in this form can be accomplished either formally; for instance, through regular meetings of designated organizational representatives, or more informally, through the ongoing but typically uncoordinated efforts of those who have a stake in network success."

The second form is called led organization–governed network. This model is more centralized and formed around a few powerful actors who make the final decision and coordinate activities. Thus, network governance becomes highly centralized and brokered with asymmetrical power. It is a hybrid model of a hierarchy and network. The third form of network governance is the network administrative organization model (NAO). The idea is that a separate administrative entity is set up specifically to govern the network and its activities. It has a centralized chain of command. Although, networks can be managed and governed to certain extent, there are also limitations of governance in loosely-coupled systems. Puustinen (2017) studied co-operation districts of health and social care as complex networks an found that management of the network may be seen as collective maneuvering in a paradoxical space of possibilities. Even if local socio-structural elements of the network can be modified, multiplexity, embeddedness and complexity make the operation of the network unpredictable and emergent.

Collaborative governance widens the scope of network governance by emphasizing the importance collaborative participatory arrangement through which non-state stakeholders such as civic organizations and citizen groups can gain access to policymaking process (Ansell & Gash 2006; McGuire 2006; Thompson et al. 2006). Ansell & Gash (2008, 544) define collaborative governance as "a governing arrangement where one or more public agencies directly engage non-state stakeholders in a collective decision-making process that is formal, consensus-oriented, and deliberative and that aims to make or implement public policy or manage public programs or assets". Interactive governance (Torfing et al. 2013) provide a new angle to governance debate by focusing on interactive forms of governance in terms of quasi-markets, partnerships, and governance networks

that seem to both challenge and transform the role of government in governing society and the economy.

Whilst there have been many advances in the field governance research, one could argue that there has been much repetition with scholars rehearsing similar arguments (especially around network governance debate) but framed from their disciplinary context and using their rhetorical lines of deviation rather than providing empirical evidence to support their arguments. Therefore, it is difficult to say whether network governance models discussed above are descriptive, normative or prescriptive. One reason for conceptual ambiguity is perhaps the fact that policymaking environments have become more complex and complicated (Geyer & Cairney 2015; Boulton et al. 2015; Stacey & Mowles 2016). In this policy nexus, people, organizations, policies and social institutions are interconnected through numerous diffuse and often hidden ties. This has increased the interdependency and unpredictability of the global system and thus made it more difficult for any single institution (state, international organizations or inter-governmental body) to control it. Bennet and Lemoine (2014) have labeled this hyperconnected and volatile operating environment as a VUCA World. This managerial acronym is derived from the words volatility, uncertainty, complexity and ambiguity.

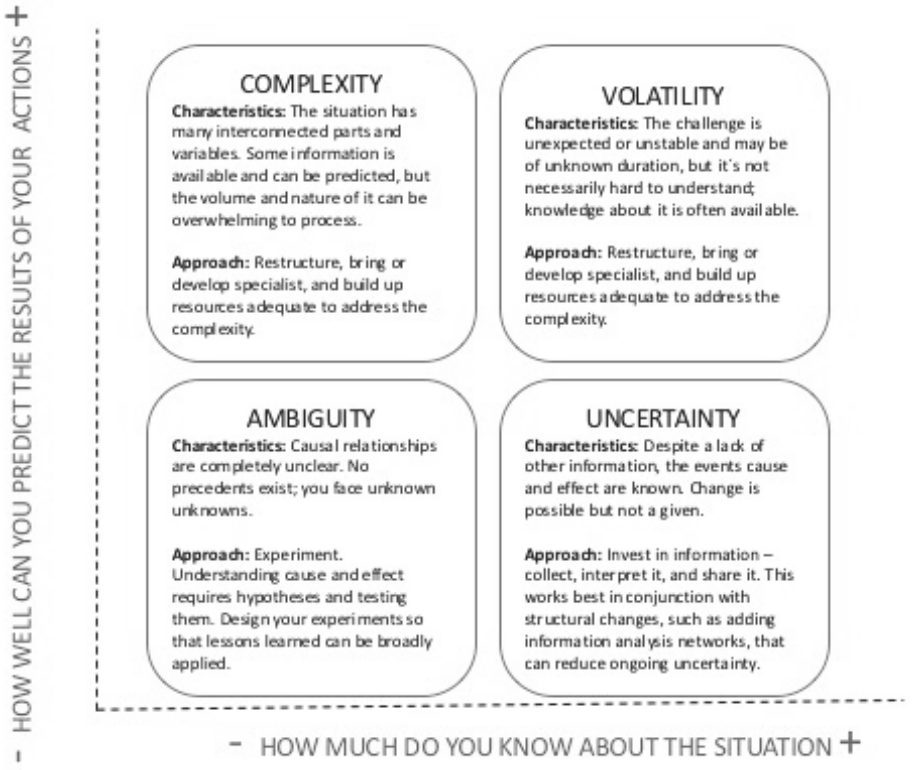

Figure 3. VUCA-dimensions and their main elements. Modified from Bennet and Lemoine (2014).

In the VUCA World (Figure 3.), the operating environment changes rapidly and is often unpredictable. Phenomena emerge rather than are created through rational planning. Ambiguity describes a situation in which the understanding of the causal relationships underlying the phenomena is completely unclear. This can be described as an *unknown unknown* (Aven 2015). In the uncertainty situation, causal relationships are known even if the information is incomplete. In a volatility situation, challenges and problems arise unexpectedly, and there is no accurate information on their duration. A complex situation arises when things are interconnected and with little information available to solve the wicked problems (Rittel & Webber 1973; Raisio et al. 2018a; Raisio et al. 2018b).

If increased complexity and unpredictability make policy analysis and decision-making difficult today, it is needless to say that it will make forecasting the policy outcomes almost impossible. The global connectedness of information, people and institutions create a fear of deterioration of trust, and an increased feeling of insecurity as the functioning of a networked society is characterized by constantly fluctuating demands and expectations as well as abrupt changes and erratic behavior. Digitalization and social media have not only made the lives of people more widely and sparsely connected but also more individualistic than before. People have easy access to information, and they can feel close to ideas, groups and people whom they have never met.

Simultaneously, the political world has become fragmented, mediatized and dramatized (Klijn & Koppenjan 2012). Policymakers have been surprised by alternative facts, fake news, disinformation, and the like. The authority of experts, traditional knowledge institutions and even scientific research, have been questioned. In addition, wicked problems (Rittel & Webber 1973; Head & Alford 2015; Raisio et al. 2018a) without simple solutions are more a rule than an exception. In communication, the challenge is that public policy lacks ways of understanding the complex interplay between rational and affective processes in the public debate (Lakoff 2008). In their cross disciplinary examination of societies as an interaction spaces Lehtimäki et al. (2020) discuss the new logics and transformation mechanisms of economic activity, citizen participation, governance and policymaking in the face of technological innovations, market-based reforms, and the threat of disconnect between citizens and policymaking. Their focus is on value co-creation in complex adaptive systems where institutions, individuals and businesses negotiate value and interests in networked relations.

This chapter has discussed the changes in governance and societal guidance from the point of view of centralized and decentralized governance structures. The shift from centralized steering to decentralized management poses major challenges for building the knowledge base behind decision-making. In the next chapter we will look more closely at the ideas behind evidence-based policymaking and above all the role and importance of evaluation in generating the evidence.

3 EVALUATION AND EVIDENCE-BASED POLICY

There is an inescapable requirement in public policy to provide evidence[1]. Many governments have been moving from opinion-based policy toward evidence-based policy, and are currently in the stage of evidence-influenced policy (Segone 2004, 27) or evidence-informed (Bowen & Zwi 2005) policy. Evidence-based Policymaking can be described as "an approach which helps people make well informed decisions about policies, programs, and projects by putting the best available evidence at the heart of policy development and implementation" (Davies 1999). It is a set of methods which "informs the processes by which policies are formulated, rather than aiming to affect the eventual goals of the policy" (Sutcliffe & Court 2005). Evidence-informed policymaking refers to the fact that the policymaking process is inherently political, and evidence translates into policy options and alternatives through different channels and mechanisms. According to Nutley et al. (2003) there are many alternative ways to define evidence-based policymaking (EBP) and practices. The content of EBP varies according to academic disciplines and evidence-based practices.

Evaluation forms a knowledge platform for evidence-based policymaking and is a critical precondition when transforming hierarchical planning systems into flexible and decentralized performance-oriented management systems (European Commission 2003). Program evaluation can be defined as a systematic operation of varying complexity involving data collection, observations and analyses, and culminating in a value judgment with regard to the quality of the program being evaluated, considered in its entirety or through one or more of its components (Pawson 2006; Pawson 2013; Patton 2011).

Evaluation means arriving at a value judgment on the basis of measures (qualitative or quantitative) considered to be valid and reliable, which compare the actual results of a program with its anticipated results. Even where evaluation is concerned with assessing intangible situations, which are difficult to measure, it must be based on data gathered in a rigorous and objective manner to be credible. Evaluation is part and parcel of a larger cultural wave consisting also audit, inspection, and quality assurance, which together constitute "a huge and unavoidable social experiment which is conspicuously cross-sectional and transnational." (Power 1997, xv). Lincoln and Guba (1986, 550) define evaluation as "a type of disciplined inquiry undertaken to determine the value (merit and/or worth) of some entity—the evaluand—such

1 Evidence can be defined as an argument or assertion backed by information. Scientific evidence refers to information gathered systematically using recognized methods or specific hierarchy of scientific methods (Cairney 2016, 3).

as treatment, program, facility, performance, and the like—in order to improve or refine the evaluand (formative evaluation) or to assess its impact (summative evaluation). Democratic evaluation can be seen as an information service to the society about the characteristics of a government program and its implementation. More detailed introduction to alternative approaches to program evaluation and their ontological and epistemological grounds are presented in Article 1. (Virtanen & Uusikylä, 2004, 77-82).

The traditional program evaluation model has been criticized for oversimplifying the nature of societal problems by defining artificial causal relationships between phenomena (Pawson 2006; Hayden 2010; Patton 2011). Another key criticism is that the premises or the theoretical assumptions underlying programs are often left undescribed and that programs are missing "a theory of change" (Mayne 2011; Montague 2009). The third crucial criticism is that program evaluation models are unable to exert sufficient control over other intervening factors affecting the phenomenon area, such as individual, regional and trend-related change factors (Virtanen & Uusikylä 2004).

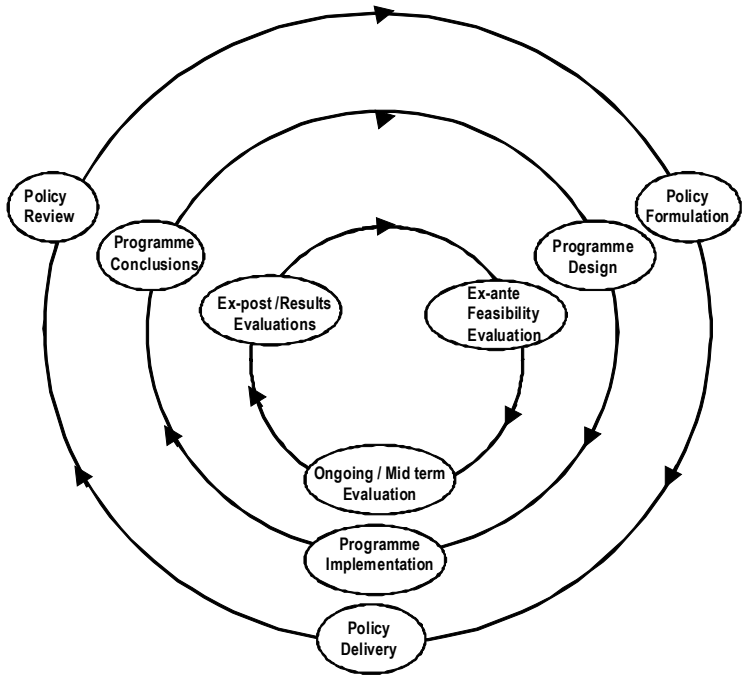

Figure 4. Evaluation linked with the policy cycle (EVALSED 2013, 8).

An alternative evaluation model can be seen in relation to the policy cycle (i.e., policy formulation, decision-making, implementation and review). Figure 4. shows the inter-linkages between policy cycle and evaluation emphases. During the policy

formulation, phase design is often referred to as *ex ante* evaluation. The purpose of the evaluation is to validate programs' or projects' assumptions and their feasibility. When policies are being implemented, evaluation and monitoring give accurate information to policymakers and program managers on how policies work (and, if not, why). This feedback is given by ongoing evaluations (either external or self-evaluations) or mid-term reviews. When the program implementation comes to its end, the evaluation focus and target are normally moved to the results, effectiveness and impacts of the program. This is called the *ex post* evaluation.

3.1 WHY EVIDENCE-BASED POLICYMAKING?

Evidence-based policymaking became popular in the late 1990s under the UK's New Labour government and its new political doctrine, the Third Way, under which there was a conscious shift away from political ideology toward evidence-guided policymaking (Sutcliffe & Court 2005). Similar government initiatives appeared elsewhere, aiming to "take forward the Modernizing government agenda," further deepening the role of evidence in policymaking (Davis et al. 2000). Simultaneously, an atmosphere of political and public skepticism was rising regarding the delivery of public services as well as a demand for greater transparency in government spending. The demand for evidence-based policymaking provided an alternative to "judgment-based professional practice" or opinion-based policy as the best way to ensure effectiveness in policymaking. The increasing role of evidence in policy has resulted from several pressures: wider public awareness, a larger research community, increased available data boosted by the technological development and ICT's, the increased intensity of international competitiveness and increased demands for government accountability (Davies et al. 2000).

As evidence-based policymaking is gaining wider support across the world, it is important to have clarity over how evidence is defined and how it can benefit policymaking. The range of what is considered to be evidence is extensive. In this essay, evidence-based policy is considered to benefit most from systematic pluralistic evidence, combining various sources and types. Good quality policymaking depends on high-quality information, derived from various sources. This includes existing research (international and domestic), expert knowledge, evaluations, and secondary information including the internet. (Davies et al. 2000, 23). These different sources of information are most beneficial when used together (Segone 2004, 29).

Policymaking can benefit from strong evidence in various ways. Firstly, evidence can reveal a new aspect of a previously unknown or hidden problem, allowing it to be addressed. Secondly, evidence informs the choice and design of policy after the policy problem has been identified. Thirdly, evidence can guide future forecasting. This is necessary in order to determine whether the policy will prove to be beneficial

in the long run. These models are important when governments set targets as they produce knowledge on whether the targets are going to be met. Fourthly, once the policies are implemented, the expected results require monitoring by the policymakers to assess whether the key results are as planned and the indicators are satisfactory. Lastly, evidence forms the basis of evaluating the impact of the policy. This is methodologically more difficult than monitoring the results, but it needs to be included in the design of the policy. (Segone 2004, 7). However, policymaking does not usually happen in one single event or contain an "explicit set of decisions" (Davies et al. 2000, 15). Rather than emerging from conscious deliberation, policy tends to develop and accrete as a result of a complex process (Weiss 1980).

Identifying a need to integrate evidence into decision-making, the UK government established the "What Works" approach and a network of independent centers aimed to support local practitioners and commissioners to

1. Undertake a systematic assessment of evidence and produce a clear and actionable synthesis of evidence,
2. Translate the evidence in a common currency enabling comparisons of effectiveness,
3. Maintain the focus on the needs and interest of the users,
4. Publish and disseminate the findings in a digestible format,
5. Identify gaps requiring capacity building or further research,
6. Support practice by advising and ensuring that projects and interventions can be evaluated effectively.[2]

There is a tendency to think that the evidence that policymakers need is always research-based. Good quality policymaking depends on information and experience derived from a variety of sources: expert knowledge, existing research, statistics, stakeholder consultation, evaluation of previous policies, or based on a synthesis of the secondary data. Thus, information can both schematic (quantitative) or tacit (qualitative). It would be naïve to assume that the policymaking process utilizes only research results or statistical data. This information and knowledge always get mixed with political values, tactical aspects and popular debates (Cairney 2016).

3.2 PUBLIC MANAGEMENT REFORM AS EVIDENCE-BASED POLICY

Despite the massive number of public-sector reform programs aimed at improved performance and accountability in all OECD countries, it is difficult to find a common

2 https://www.gov.uk/guidance/what-works-network

understanding of or clear empirical evidence on the success of these reforms. Pollit and Bouckaert (2003a) have scrutinized the results of public-sector reforms in a comparative setting and find only modest evidence of improved economy, efficiency, processes, effectiveness and systemic changes. Even in areas such as savings and improved processes where some positive development can be found, it is questionable whether this is a result of managerial reforms or is dependent on other factors.

Evidence-based policy is seen to address the key elements of the policymaking process poorly, as it is unable to respond to the inherently complex, ideological and context-dependent policy environment. Additionally, it has been criticized for de-politicizing policymaking, which is the main reason that it has not been implemented in France (Oxfam 2013). However, evidence-based policymaking does not necessarily have to assume a linear causal relationship between evidence and outcome, but that evidence can influence policymaking through a more cyclical, iterative approach, by renewing evidence throughout the project and reacting to the ever changing context. This approach requires more capacity from the policymakers as well as more resources that are often lacking in developing countries. The inherently political nature of policymaking and evidence should be accepted, and, therefore, the evidence-base should be as broad as possible, presenting information from various sources.

Under scrutiny, these "obvious facts" reveal a wider ontological minefield that is related to societal values, the role of democratic institutions as allocators of societal values and the difficulty in separating correct information from the incorrect.

Critical questions, such as the following, may be brought up with justification:

1) What are the ideological connections between performance management and evidence-based decision-making and program planning?
2) Why are these discourses gaining traction while people are calling for value-based management and decision-making?
3) What is the societal value base for these models?
4) What parties are speaking for evidence-based operational culture, and whose interests do they ultimately serve?
5) On what type of theory of change are these models based, and how do they support citizens' participation and ability to influence political decision-making?

It would be too simplistic to assume that policymakers make decisions based on perfect information. Bounded rationality together with political values or simply party political or organizational tactics block good attempts to provide evidence (either *ex ante* or *ex post*) to policymaking processes. Sometimes it is in the self-interest to policymakers to "leave political airspace" untouched. Cairney (2016, 5) emphasizes a need to understand the psychology of policymaking. He states that

policymakers have to make important decisions in the face uncertainty, which is based on limited information, ambiguity, which is based on the fact that there are many ways to understand a policy problem.

The field of organizational theory has had long lasting criticism of the idealism of the rational planning model. Simon (1957) has been the most quoted scholar on the topic, defining the limits of rationality. According to Simon, bounded rationality has to be considered when analyzing how individuals make decisions. In most cases, their rationality is limited by the tractability of the decision problem, the cognitive limitations of their minds and the time available to make the decision. For Simon, organizational decision-making is, therefore, characterized by satisficing rather than maximizing behavior.

Many other critics of the rational model emphasize its inadequate representation of the policymaking process in practice. For Lindblom (1959), policymaking neither can nor ought to follow the stages of the rational model. His alternative interpretation has been labeled as disjointed incrementalism or the science of muddling through. Braybrook and Lindblom (1963) argue that decisions are not made as a movement toward predetermined goals, but instead, the process is more piecemeal. Problems are tackled when they emerge in the process of muddling through. Later, Amitai Etzioni introduced his synthesis of rationalist planning theory and incremental theory. The main argument of his mixed scanning approach was that fundamental decisions should adopt a rationalist approach where more routine decisions can be made in a more incremental fashion (Etzioni 1967).

Some other models of policymaking have shared this idea of bounded rationality (Simon 1957, Lindblom 1959) or even irrationality (Brunsson 1987) of decision-making. Garbage-can theorists (Cohen & March 1974; March & Olsen 1976) have noted that policy solutions do not necessarily arise from systematic problem identification and analysis. Instead, pre-existing solutions can result in a search for problems to which they can become attached. The model symbolizes the decision-making situation as a garbage can that participants chaotically dump problems and solutions into as they are generated. The model portrays problems, solutions and decision-makers as three independent streams that are each generated separately and flow disconnectedly from one another. As Radej (2015, 9) defines the problem, "Rigorously non-complex approaches to studying social processes, like micro or/and macro models in mainstream sociology and economics, may have problematic construct validity when their theories fail to reproduce a logic on the middle level of considerations, which is thoroughly complex but also most intrinsically social."

Granovetter (1985) offers an alternative, i.e. a relational approach to the decision-making process. Granovetter (1985, 504) states that "most behavior is closely embedded in networks of interpersonal relations and that such an argument avoids

the extremes of under- and oversocialized views of human action."[3] Granovetter comes quite close to some of the transaction cost theorists (e.g. Ouchi and Williamson) who have raised a similar kind of criticism against the under-socialized neoclassical economic theory. Williamson (1979) identifies the critical dimensions of characterizing a transaction and links these to the institutional governance structures of transactions. The principal dimensions describing a transaction are uncertainty, frequency of exchange and the degree to which investment are transaction-specific. According to Williamson (ibid.) the efficient organization of economic activity entails the matching of governance structures with these transactional attributes in a discriminating way. He concludes that non-specific transactions are efficiently organized by markets, while recurrent transaction-specific exchanges are more efficiently governed internally.

According to Ouchi (1980), evaluating organizations according to an efficiency criterion would make it possible to predict the form organizations will take under certain conditions. Organization theory has not developed such a criterion because it has lacked a conceptual scheme capable of describing organizational efficiency in sufficiently microscopic terms. In spite of the apparent parallels between Granovetter's thoughts and the transaction cost theorists, he still argues that "Williamson treats these examples as exceptions and also fails to appreciate the extent to which the dyadic relations he describes are themselves embedded in broader system of social relations. I argue that the anonymous market of neoclassical models is virtually nonexistent in economic life and that transactions of all kinds are rife the social connections described" (Granovetter 1985, 495).

Bouckaert and Halligan (2008) have referred to governance reforms as a sort of perpetual-motion machine used to develop solutions to various problems in public policy, which in time create new problems in public policy requiring resolution with a different method (also see Bouckaert & Van Dooren 2009; Van Thiel 2016). Van de Walle (2016) has stated that public governance reform policy typically proceeds through contradictory objectives, which is reflected as illogical decision-making regarding governance practices and structures. Bearing such observations in mind, the approach to public governance reforms should be consistent and incremental. It is essential to take traditions of political-administrative culture into account, but reform policies should nevertheless be aimed at the achievement of a long-term vision for the development of governance policy (Virtanen et. al 2016).

The systems seem to be biased in favor of a macro view of reality, which is a serious shortcoming in the face of social complexity (Radej 2015, 23). Let us take

3 Granovetter refers to Denis Wrong (1961) who complained about an "oversocialized conception of man in modern sociology" by which he meant overemphasis of consensually developed systems of norms and values, internalized through socialization. Here one can perhaps see reflections from Parsons (1977) functional theory that dominated the academic discussion of that time on social action.

the theory of governance as an example to demonstrate the problems providing evidence to the existence of causal relations in the process of macro institutional change, i.e., systems change. In the NPM literature, it is often implicitly assumed that the introduction of a new administrative model or reform leads to a certain expected change in society, for example, the improvement of efficiency or innovativeness (see Figure 5.).

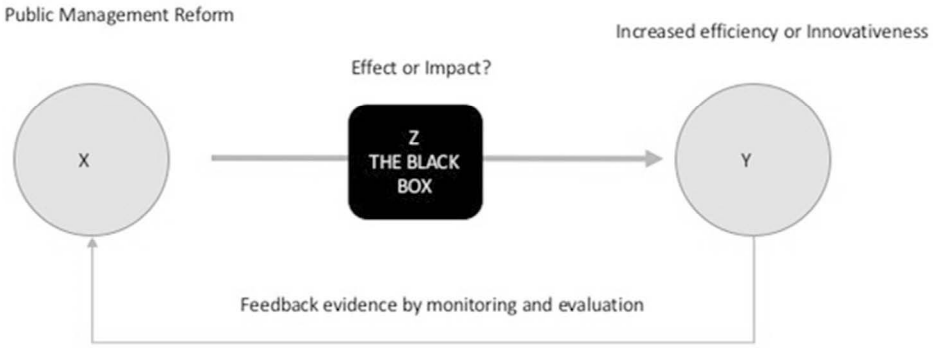

Figure 5. Macro-proposition: public management reform increases efficiency or innovativeness (source: the author).

Even if this better government assumption is (explicitly or implicitly) behind most public administration reforms (see OECD 1998), it is relatively hard to prove this causal linkage between the reform introduced and policy outcomes observed. The major problem for explanations of system behavior based on actions and orientations at the level below that of the system is moving from the lower level to the system level.[4] Under this approach, the public sector as market and the economy itself can be viewed as an outcome of the interactions of individuals, firms and the state over time, following Karl Polanyi's (1957) notion of the 'embeddedness' of the economy in society and culture. The use of the example presented in Figure 5 raises further questions about what kind of proposition was demonstrated, and, in particular, what unit or units were involved in the proposition. Should the proposition be specified at the individual level? If so, then the proposition in Figure 5 must be revised. The single proposition breaks into three: (i) one with an independent variable characterizing the individual; (ii) a second with both independent and dependent variable characterizing the individual; and (iii) a third with the independent variable characterizing the individual and dependent variable charactering the system.

4 In economics, for example, there is microeconomic theory and there is macroeconomic theory; and one of the central deficiencies in economic theory is the weakness of the linkage between them, a weakness papered over with the idea of "aggregation" and with ubiquitous concept in macroeconomic theory, that of the representative agent (Coleman 1990, 6-7).

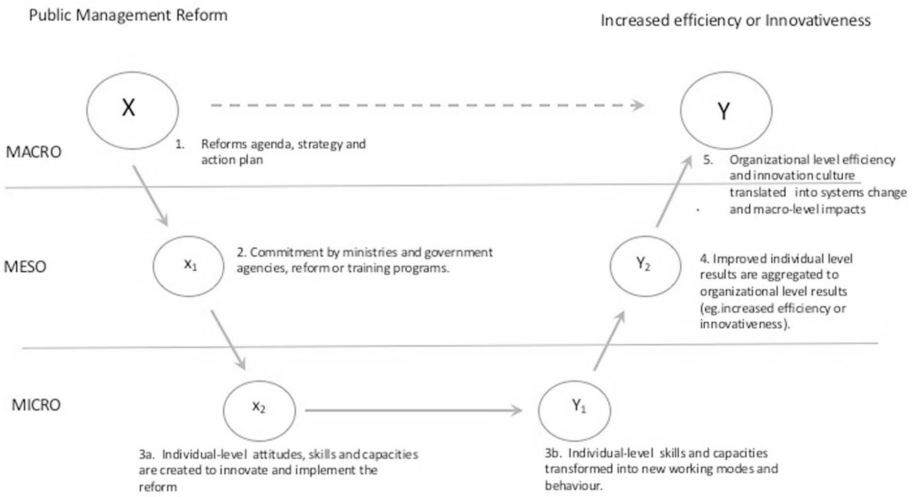

Figure 6. Macro- and micro-level propositions: effects of public administration reform on a public management system. Modified from Coleman (1990, 8).

Figure 6 shows a way of diagramming such multilevel systems of propositions. The model is based on Coleman's idea of bridging the macro–micro gap (1990).[5] The upper horizontal arrow represents our baseline macro-proposition. There are three connected arrows. The first begins from the macro-level proposition and goes down to the lower level (individual level). This arrow (i) represents the transformation function that converges with the public administration reform objectives into individual or organizational, administrative values (such as customer orientation and sound financial management principles). The second lower-level arrow (ii) represents the function that turns administrative values into action (e.g., managers actually start acting according to customers' needs or cost efficiency). The third arrow (iii) is the most important because it moves back up from the individual level to the system level. In concrete terms, it verifies the aggregation of changes taking place at the individual level behavior and resulting in system level changes (e.g., increasing government's efficiency, improving management culture or adopting new innovative administrative culture).

Changes in administrative culture are often urgently needed to support the implementation of public management reforms or performance management systems. Pollit and Bouckaert (2003b) conclude that the real results of administrative

5 The Coleman model has been criticized for oversimplifying the macro-micro conversion logic and especially the causal assumptions behind it (see e.g. Hedström & Svedberg 1996; Hedström & Swedberg 1998; Abell 2000), or neglecting the meso-level intermediaries such as organizations (see Jepperson & Meyer 2011; Ylikoski 2016) or dynamic changes taking place in time (Ylikoski 2016). For more detailed analysis of the shortcomings and alternative models see Ylikoski (2016).

reforms are seldom properly evaluated and that the main agents behind the reform rarely follow the rules of their game themselves. The performance management practices often manifest themselves most clearly in symbolic action and vague performance rhetoric, which does not sufficiently help public agencies and programs in reflexive correction of their own action.

In order to avoid the aforementioned problems, the latest innovations in public management emphasize balanced measurement frameworks and comprehensive evaluation systems. Although they have different origins, there is a common denominator to these various approaches: they all try to map the causal relations between far-reaching goals, short-term objectives and the means needed for their achievement.

3.3 HOW EVALUATIONS SUPPORT EVIDENCE-BASED POLICYMAKING?

Evaluation is a critical component of policymaking, programs and projects at all levels of government. In general terms, evaluation allows for the informed design and modification of policies and programs to increase their effectiveness and efficiency. With accurate and reliable information, evaluation provides program management teams and other interested parties with the means to learn from experience, including the experience of others, and to improve service delivery. It serves the dual function of providing a basis for improving the quality of policy and programming, and a means to verify achievements against intended results or unintended consequences (positive or negative).

Evaluation should provide answers to the two-sided question: "Are we doing the right things, and are we doing things the right way?" With answers in the affirmative or with action plans to respond to areas of weakness, evaluation nurtures political and financial support of appropriate policies and helps governments build a sound knowledge base. Thus, evaluation can have a strong advocacy role as well as enhance the sophistication and quality of institutional performance[6].

Expectations concerning the role of evaluators vary between different evaluation cultures and the paradigms applied. The evaluator is, in turn, expected to be a neutral judge, facilitator, provider of accountability and sometimes even a problem or conflict solver (Vedung 1997). The level of institutionalization varies from country to country. So does also the *modus operandi*. By applying a neo-institutional approach, Ahonen (2015) shows how ambiguous the institutionalized practices in evaluation are and how imprecise the agency of the actors that carry out or commission evaluations or

6 For a comprehensive overview on alternative evaluation approaches, models and methods see Fox et al. (2017).

utilize the evaluation results. This is likely to create disintegration, overlapping or conflictual practices and even blur the overall policy understanding and learning. Program evaluation can be defined as a systematic operation of varying complexity involving data collection, observations and analyses, and culminating in a value judgment with regard to the quality of the program being evaluated and considered in its entirety or through one or more of its components.

Modern evaluation operations are an essential part of transparent and open public administration. However, evaluation is no mere technical instrument for monitoring results and effectiveness. Two wider social discourses affect evaluation: the requirement for evidence-based decision-making and performance management in public administration. Both of these evaluation premises are usually approached via a technical–rational cognitive interest. Decisions must be well-prepared and based on researched information, and the public administration must be able to demonstrate to the citizens what has been accomplished with tax-payers' money.

Practical problems pertaining to the selection of the evaluation framework or model are often encountered in evidence-based decision-making—again, something that evaluation attempts to promote. The model applied largely determines the type of evidence supplied to decision-makers by the evaluation and the kind of preconditions set for the evaluator by the selected models. The following is a practical example that attempts to demonstrate the type of evaluation premises that can be applied in order to demonstrate the results and effectiveness of social-policy programs or projects. Models range from a simple program model to more complex and comprehensive systemic models.

Evaluation means arriving at a value judgment on the basis of measures (qualitative or quantitative) considered to be valid and reliable that compare the actual results of a program with its anticipated results. Even where evaluation is concerned with assessing intangible situations, which are difficult to measure, it must be based on data gathered in a rigorous and objective manner to be credible.

The objective of the experimental evaluation approach is to clarify as completely as possible the causality (causal connections) of the intervention being evaluated, while also controlling the impact of intervening variables on the change observed in the evaluation. This can also be used to test the underlying assumptions of a program or project, and the explicitly applied intervention logic, i.e. whether the program generates the desired results and effects (cf. e.g. Campbell & Russo 1999; Virtanen 2007; Mohr 1995). Randomized controlled trials or randomized impact evaluations have been introduced as rigorous models for evaluating the net impact of a policy or a program. Randomized controls allow individuals to be placed randomly into two groups—those that receive the intervention and those that do not. This allows the researcher to determine program impact by comparing means of the outcome variable between the treatment and the control group. Quasi-experimental (non-random) methods can be used to construct controls when it is not possible to obtain

treatment and comparison groups through experimental design. With constructed controls, individuals to whom the intervention is applied (the treatment group) are matched with an "equivalent" group from which the intervention is withheld and the average value of the outcome indicator for the target population is compared with the average of that for the constructed control (White & Sabarwal 2014).

These RCTs are a type of impact evaluation that uses randomized access to social programs as a means of limiting bias and generating an internally valid impact estimate. Randomized controlled trials give answers to questions such as what is the causal effect of training or education on employability? What is the causal effect of incubation activities on the growth of the start-up company? The problem with these questions is that one should be able to answer counterfactual questions such as how would start-ups that participated in a program have succeeded in the absence of the program? How would those who were not exposed to the program have fared in the presence of the program? In order to respond to these questions, there needs to be a test group and a comparison or control group. Experimental or quasi-experimental designs are used to evaluate the net effect or impact of policy intervention.

Counterfactual evaluation design is one way to understand whether policies really work and under what circumstances. In the evaluation field, the counterfactual design originates from an article by Campbell and Stanley (1966) in which they sketch alternative impact evaluation designs and their quality standards.

According to the Work Bank (2009), "Estimation methods broadly follow evaluation designs, with the determination of the counterfactual being the core of the evaluation design. Different ways to establish controls under different evaluation designs determine the methods used to measure the counterfactual":

- Randomized controls allow individuals to be placed randomly into two groups—those that receive the intervention and those that do not. This allows the researcher to determine program impact by comparing the means of the outcome variable between the treatment and the control group.

- Quasi-experimental (non-random) methods can be used to construct controls when it is not possible to obtain treatment and comparison groups through experimental design. With constructed controls, individuals to whom the intervention is applied (the treatment group) are matched with an "equivalent" group from which the intervention is withheld. The average value of the outcome indicator for the target population is compared to the average of that for the constructed control.

- Another non-random method of obtaining control involves reflexive comparisons where participants who receive the intervention are compared to themselves

before and after receiving the intervention. Statistical controls or instrumental variables are used in cases when participants and non-participants are compared, controlling for other characteristics which may be statistically different between the two groups).

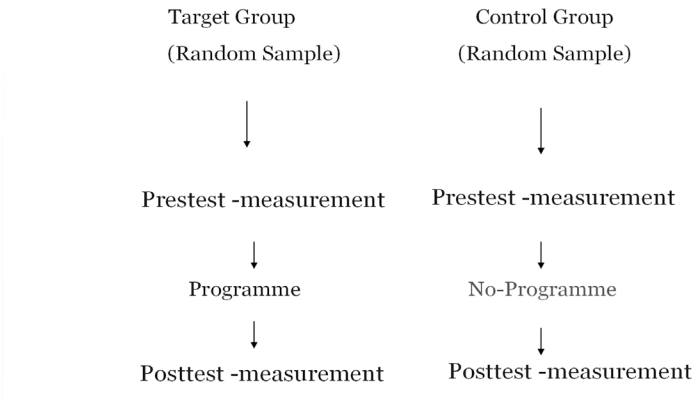

Figure 7. Quasi-experimental evaluation design (source: the author).

Figure 7 describes the quasi-experimental evaluation design. According to Kirk (2013, 24), "an experimental design identifies the independent, dependent, and nuisance variables and indicates the way in which the randomization and statistical aspects of an experiment are to be carried out. The primary goal of an experimental design is to establish a causal connection between the independent and dependent variables. A secondary goal is to extract the maximum amount of information with the minimum expenditure of resources." In most real-world social experiments, it is not possible (due to ethical reasons) for participants to be randomly assigned to treatment levels. Therefore, quasi-experimental designs are more frequently applied. If the research has all of the features of an experiment except random assignment, it is called a quasi-experiment.

A counterfactual impact evaluation gives some suggestions for the potential impact of a policy or program, but its limitations are inevitable. Firstly, the approach applied in the counterfactual impact evaluation is rather narrow and limited. Howard White an advocate of theory-based[7] impact evaluation, summarizes the criticism as follows:

7 Theory-based evaluation is an approach in which attention is paid to theories of policymakers, programme managers or other stakeholders, i.e. collections of assumptions, and hypotheses - empirically testable - that are logically linked together.

> Criticisms of reporting an average treatment effect should not be overstated. Heterogeneity matters, as does understanding the context in which a particular impact has occurred. But it will rarely be the case that the average treatment effect (usually both the treatment of the treated and the intention to treat) is not of interest. Indeed, it is very likely to be the main parameter of interest. It would be misleading to report significance, or not, a particular sub-group if the average treatment effect had the opposite sign. Moreover, the average treatment effect is the basis for cost effectiveness calculations. (White 2009, 17).

To sum up, "how much difference does a policy or program make" is a relevant evaluation question from the point of view of the evidence-based policymaking. It, however, often fails to explain the causal mechanism behind the observed change. The ultimate objective in asking these questions is to learn whether the intervention works and which interventions produce the desired effect. Or, as seen from a different perspective, to what extent are the observed changes truly caused by the intervention.

Often quoted is causal contribution analysis (Mayne 2011; Leeuw 2003) that aims to demonstrate whether or not the evaluated intervention is one of the causes of observed change. Contribution analysis relies upon chains of logical arguments that are verified through careful field work. Rigor in causal contribution analysis involves systematically identifying and investigating alternative explanations for observed impacts.

Contribution analysis (e.g. Mayne 2011; Montague 2009; Wimbush et al. 2012) has been offered as a new way of discerning the theory of change and causal chains underlying societal programs and of providing a critical evaluation of a given program's contribution (instead of the traditional attribution) to solving societal problems. The objectives of this model also include separating the program's effects from the perspectives of the various target groups and identifying the roles and responsibilities of various actors with regard to program implementation.

Contribution analysis is based on an impact chain approach (logic model) presented in relation to a simple program evaluation model, although it attempts to elaborate on this by looking explicitly at the assumptions underlying a change, highlighting critical content-related questions and risk factors and conducting empirical measurement of the effects that factors outside the program may have on the program's results and repeatability.

Perhaps the most visible approach is realist evaluation (Pawson and Tilley 1997; Pawson 2006) that stresses the epistemological differences of proposing a different understanding of causality based on a "generative" notion centered on the identification of causal mechanisms rather than a mere "successionist" view, typical of the counterfactual approach. The basic idea of a realist evaluation is that different

contexts may yield different reactions to the same intervention and putting in place alternative mechanisms may produce different results. According to Pawson:

> The nature of causality in social programs is such that any synthesis of evidence on whether they work will need to investigate how they work. This requires unearthing information on mechanisms, context and outcomes. The central quest is to understand the conditions of program efficacy and this will involve the synthesis in investigating for whom, in what circumstances, and in what respect a family of program works. (Pawson 2006, 25)

The existence of multilevel forms of governance and of various players at the different levels of the steering system (in this case: global, European, national, regional and local) make it even more difficult for decision-makers to see the "big picture" and to make informed decisions that will really have an impact. All of this has altered the dynamics of policymaking and set new restrictions on the credibility of traditional democratic governance models. Traditional evaluation models have been inefficient to cope with increasing complexity and systemic development in global politics and development aid.

The most obvious difficulties seem to be

1) Suboptimal performance orientation. This means that each agency tends to define its performance targets only from its own narrow perspective, which, at the aggregate level, leads to suboptimal results.
2) Attribution. Government organizations are unable to demonstrate their contribution to overall results (e.g., effectiveness). This causes problems in terms of accountability.
3) Invalid performance indicators. Use of performance indicators that do not capture the essential substance of the verbally expressed strategic goals.
4) Insufficient steering. Ministries lack the steering capacity and are not systematically reviewing the achievements of performance targets used.
5) Uniqueness delusion. Public agencies claim that their activities are so unique and specific that it is hard to find valid indicators to measure their performance.
6) Reporting. Lack of consistent and informative performance reporting.
7) Responsibility and accountability. Government agencies are not being held responsible for their performance.
8) Lack of incentives and reward mechanisms. Since valid performance measures are lacking or biased, evidence of their use as a basis of rewarding schemes both on an individual and on an organizational level remains scarce. (Uusikylä & Valovirta 2007, 401–402)

Typically, the evaluations of public interventions (especially in development policy) are based on linear input–output logic models. However, there has been an increasing criticism that the results-based management model (RBM) and logic model approaches to evaluation are not sufficient tools for evaluating public interventions in complex settings (e.g. Patton 2011; Pawson 2013). Rationalistic (meaning synoptic) planning frameworks, which embed the causal logic behind actions (from inputs to outcomes and impacts) have been developed and used extensively, especially in the field of international development aid. The logical framework (or logframe[8]) approach has been the mainstream tool for planning aid interventions, both at program and project levels. Logframes provide a simple and useful planning tool and the basis for evaluating projects or programs. Logic models are grounded in theory-based evaluation and provide a coherent theory on how activities are intended to generate the results.

Evaluation literature highlights that a program may fail for two reasons: either the program has failed to put the intended activities into operation (implementation failure), or the activities have failed to bring about the desired effects (theory failure). When the logic of the project or program is flawed, or the activities and outputs do not generate the desired results, the project or program is likely to fail even if the process of implementation is successful.

However, in an era of fast-paced change, it is not always possible to predict changes in the economic, social and environmental contexts that impact the success of the program objectives. Too often the inflexibility of the logframe approach can limit staff capacity to adapt to emergent trends by holding them accountable for predicted cause and effect rather than accountable for the ability to learn from the use of rigorous evidence analysis in implementation, to adapt to changing circumstances and to understand the conversion mechanisms that translate inputs into outputs, outcomes and impacts (Virtanen & Uusikylä 2004).

Practical problems pertaining to the selection of the evaluation framework or model are often encountered in evidence-based decision-making—again, something that evaluation attempts to promote. The model applied largely determines the type of evidence supplied to decision-makers by the evaluation and the kind of preconditions set for the evaluator by the selected models. The following is a practical example that attempts to demonstrate the type of evaluation premises that can be applied in order to demonstrate the results and effectiveness of social-policy programs or projects. Models range from a simple program model to more complex and comprehensive systemic models.

8 As a methodology, the Logical Framework Approach (LFA) is a systematic, visual approach to designing, executing and assessing projects which encourages users to consider the relationships between available resources, planned activities, and desired changes or results. https://www.betterevaluation.org/evaluation-options/logframe

The interdependence, complexity and multilevel nature of societal phenomena have increased dramatically over the last decade (Easley & Kleinberg 2010; Byrne 2005). The primary reasons for this are developing information technology, changes in the global network economy, a multilevel decision-making environment (global, national, regional, local) and the increasing rate of change.

This causes problems not only with regard to compartmentalized and sector-based decision-making systems (Uusikylä 2013) but also with the evaluation of political-administrative decisions and societal phenomena. If we genuinely wish to understand wicked problems and the overall impact of political decisions, we must replace the current program and project evaluation models with alternative evaluation frameworks. The starting points for these models are strong interdependence between phenomena and nonlinearity and an attempt toward holistic interpretation in which the whole is greater than the sum of its parts (Byrne 2005, 3). On a small scale, this is about generating a new type of evaluation paradigm (cf. Kuhn 1996). Before we outline the basic elements of this new paradigm, we should look at the assumptions of a traditional, program theory-based evaluation model through a practical example case.

The existing paradigm of the OECD/DAC, the EU and most international organization follows the rationalistic program theory models in which evaluation focuses strongly on program inputs, outputs, results and impacts (the so-called logic model). The logic model has a strong relationship to evaluation approaches that have been grouped under the label of theory-based evaluation (Weiss 1997; Weiss 1998; Pawson 2013). This tradition tries to structure evaluation to reflect a coherent theory of how activities are intended to generate results. Although, the limitations of this model are well-known and also expressed in the ECPE-guidelines and handbooks (e.g., EVALSED 2013), the model has been applied in a rather mechanic way. The delivery mechanisms are not analyzed and understood sufficiently (the so-called black-box syndrome), and the contextual factors that shape policies are not given enough attention. The linear evaluation model or logic model applied by the European Commission is presented in Figure 7 (for the logical model and its limitations, see Uusikylä & Valovirta (2007)). Godenhjelm (2016) concludes in his academic dissertation that the growing use of project-driven organizations in the public sector has significant consequences, as well as showing that the expected advantages of project organizations are related to the rationalistic ideals (see below), but also that temporality as such poses challenges to permanent administrative structures. There is a threat that projectification little by little becomes the new normal in public service delivery (Sjöblom et al. 2013; Vento 2017).

Figure 8 shows that the chain of cause and effect running from the input to the output, to the outcome and finally to the impact. The first two can be defined as follows:

Input: The input *is* the intervention. It is usually measured in cost terms (e.g., the amount of resources allocated or number of man-days resourced), but it can be measured in other units, such as the hours of advisory support to develop entrepreneurship or hours of training to enhance worker or management skills.

Output: The intervention generates outputs via a set of activities or processes. For a job creation project, it is the increased number of jobs in supported firms. However, it could be the number of firms completing a training course or the number of individual being trained on a project.

In general, the basic problem in evaluation (whether *ex ante* or *ex post*) is that not all of the outputs will be attributable to the intervention. For example, some of the firms creating jobs may have created the jobs anyway, irrespective of the project. Further, there may be indirect effects that occur in other firms. For example, supported firms may put other non-assisted firms in the local area out of business. In this way, it is also important to distinguish between the outcome and impact.

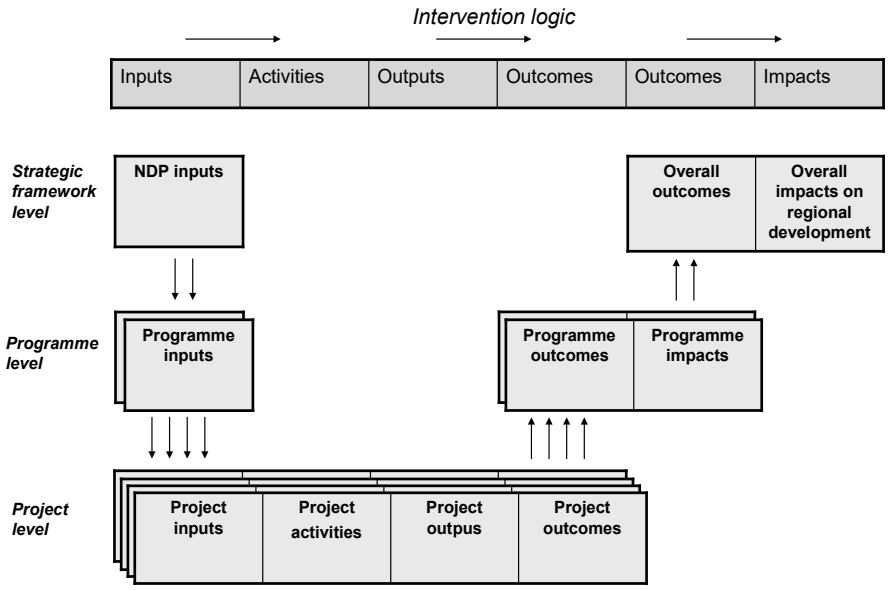

Figure 8. A linear evaluation model applied by the European Commission (EVALSED 2004).

Outcome: The outcome (or what is sometimes known as the result) is the short-run effect experienced by the agents or markets directly affected by the strategy, for example, over the first year or so. In the case of the job creation project, it is the jobs in supported firms that would not have been created without the project.

> **Impact:** The impact is the longer-run, economy-wide or societal effect experienced by all actors within the local area, for example, over a period of three to six years. This allows for effects of the job creation project on other firms, including feedback effects on those firms that were supported by the project.

Logic models are most useful when sketching and planning programs or projects. However, their use in evaluation is problematic both for methodological and practical reasons. Linear simple systems are based on the assumption that an increase in the size of an input to the system gives a proportional increase in the size of the output. In this reasoning, one event must lead to another in either a deterministic (absolute) way or in a stochastic (probable) way (Andersson 2003) so that a future state can be precisely predicted from a previous state. Methodological flaws relate to the lack of understanding of social mechanisms that produce outcomes (see, e.g., Hedström & Swedberg 998) and understanding of the contextual factors that work under certain circumstances but might not produce the desired effect under others (Pawson & Tilley 1997). Figure 9 shows a logic of simple causal mechanism. Practical problems have to do with difficulties in finding reliable and valid monitoring and evaluation indicators, aggregating data from outputs to outcomes and long term impacts, proving the attribution and net effect of particular program interventions and finally utilizing evaluation findings in reformulating policies.

Evaluations tend to label, re-label and describe rather than explain why certain policies work while others do not. One of the main problems behind this and the mechanistic causal-effect models is that they neglect the fact that action always takes place in situations, and, thus, their success depends upon the way the action is performed by specific actors in specific situations (as the realistic school of evaluation has emphasized). In a black-box explanation, the link between input and output, or between *explanans* and *explanandum,* is assumed to be devoid of structure, and the explanatory mechanism is simply considered to be a regression coefficient linking I and O (Hedström & Swedberg 1998, 9–10).

Figure 9. A causal influence chart (adopted from Kaufmann 1987: 10–11).

An alternative approach (the so-called mechanism-based explanation) seeks to find a causal agent that is assumed to have generated the relationship between the entities being observed. Kaufman (1987) highlighted that fact that even before the rise of realistic evaluation tradition approach, originally heavily coined by Pawson and

Tilley, by the following two chains of logic: by concentrating on mechanisms, one can capture the dynamic aspect of scientific explanation; yet, it is important to remember that causal explanations must be distinguished from true causal statements (Elster 1989, 4). Figure 10 illustrates the fact that traditional cause–effect logic totally undermines the fact that the effects of a program are always caused by real actors rather than constructed ideal actors. A program always achieves influence through societal actors that base their courses of action on interpretations of the expected logic of a program—that is, actions are always socially constructed rather than objectively derived from abstract program logic. Programs always interfere with an intervention field of social actors whose reactions produce final outcomes or effects.

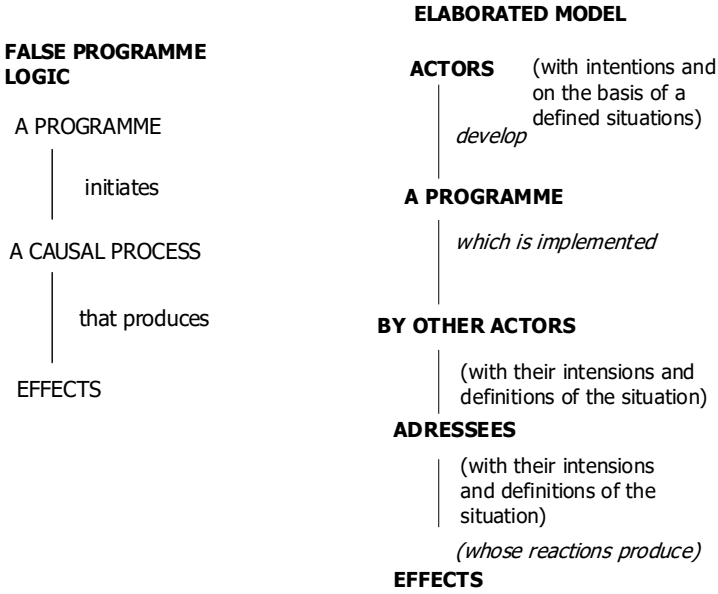

Figure 10. The true and false social intervention logic of a program (adopted from Kaufmann 1987: 10–11).

The elaborated model (i.e., the mechanism-based explanation) begins with an understanding that the programs themselves are socially constructed, i.e., certain agents (with intentions and their best understanding of societal needs) develop programs, which thereafter are implemented by other actors who might—and most often do— have their own interpretations of program logic and its situational validation. Secondly, after that, there is a third set of actors who try to adopt the expected goals and objectives (based on their own subjective understanding) and finally implement the program. Following this logic of program implementation constructed by multiple sets of actors, it seems to be self-evident that we are dealing with several rather than one single interpretation of a program. Therefore, we need more precise methodology to understand how social programs turn into real

outcomes and impacts. In order to do this, we also have to tackle the so-called agency problem or problem of combining micro-level processes to macro-level policies. In the main stream program evaluation models, there is an imminent gap existing between program-level goals and project-level results. This is a classical problem of aggregation: program-level results cannot simply be aggregated from project-level results, although most of the program evaluators and commissioners of these evaluations tend to believe so. This is quite close to what Fischer (1995) called mapping out situational validity in program evaluation.

This chapter has examined the theoretical and pragmatic roots of evidence-based policymaking and considered the importance of evaluation in generating the underlying knowledge. In addition, the chapter has discussed the shortcomings of linear program models and the problem of causal inference with reference to the articles of this dissertation. The review has been somewhat indigenous, deterministic and reductionist. To understand the problems associated with the practical application of linear programming, the phenomenon need to be viewed from a broader viewpoint of the open system. Complexity theory states that physical, biological and social systems can be divided into three categories. Ordered systems are predictable and linear: causality moves in a single direc- tion. Chaotic systems are subject to constant, mathematically unpredictable change and complex systems are defined by having many autonomous actors that have multiple interactions with each other. The actors are interdependent (Kleinfeld 2015, 18). The open systems interpretation starts with the idea that forms, patterns, and institutions emerge rather than be created and are then constantly challenged and potentially invaded by particular events, variations, decisions, shocks, and so take place at particular times (Boulton et al. 2015, 29; Prigone et al. 1977).

Opening causal chains and the "black box" described above is important from the point of view of utilizing evaluation information. Ultimately, it is about understanding the nature of knowledge and, through that, about the legitimacy of the utilization of evaluation results. Marketta Rajavaara (2007, 50-51) argues that due to New Public Management and neo-liberal political culture "the concept of effectiveness made a breakthrough, and new accountabilities – with their knowledge tools for performance measurement and auditing – and evidence-based styles of reasoning became more dominant in the ruling of the welfare state.

The use of evaluation information and has long been the subject of debate among evaluation scholars (e.g. Weiss 1998b; Patton 1997; Dahler- Larsen 1998). Michael Q. Patton has emphasized that evaluations need to be both utilization-focused and highly dissemination-oriented (Patton 1997). According to him without these elements evaluations should not be carried out at all. For other scholars like Erik Ahlbæk (1996) and Peter Dahler-Larsen (1998), dissemination is not so prominent. Ahlbæk is quite pessimistic, saying that "evaluation utilization research has generally been theoretically underdeveloped, and models of evaluation utilization have

been underspecified" (Ahlbæk, 1996, 2). One of the central findings in Dahler-Larsen's study (1998, 64–5) was that in evaluation research, the utilization of single evaluations is usually quite limited. According to him, the growth, diffusion and institutionalization of evaluation procedures have taken place before the most fundamental problem concerning the utilization of evaluations – the dissemination preconditions and effects – has been resolved. Valovirta (2002) agues that evaluation utilization discourse neglects the argumentative perspective. According to Valovirta rather than regarding evaluative information as indisputable knowledge, it should be viewed as a collection of arguments, which can be debated, accepted and disputed. To follow this line of argumentation, evaluations consist of different kinds of statements, which become matters of individual interpretation, collective argumentation and decision making in interactional contexts.

Carol Weiss (1998a) proposes that there are four main types of use describing how officials in executive agencies and legislatures use research and evaluation information:

1. Guidance for action where results are used as direction for changes in program and policy.
2. Reinforcement of prior beliefs is when evaluation findings affirm what people already know and believe about the pitfalls in the program. It bolsters the confidence of those who want to press for needed change.
3. Mobilization of support occurs when evaluation findings are used to mobilize a coalition of support for change. The findings become ammunition in organizational debates.
4. Enlightenment is when evaluation results are used for a general increase in understanding and people learn more about what happens in the program and afterwards. They gain a better idea of program strengths, fault lines and opportunities for improvement.

Fleischer & Christie (2009) have found five types of use evaluation findings and results. The first three are primarily related to the use of findings per se, i.e. instrumental usage, when decision makers use the evaluation findings to modify policies or programs (e.g. Green 1988). Utilization can also be conceptual. This takes place when the evaluation findings help program managers or policymakers understand the problems or root-causes in a new way (Weiss 1979). Third instrumental use according to Fleischer & Christie relates to wider usage of evaluations by citizens, stakeholders or media. In addition to the three types of use referred above, Patton (1997) describes a fourth type of use which is process use. Process use is defined as "cognitive, behavioral, program and organizational changes resulting from engagement in the evaluation process and learning to think evaluatively" (Patton, 2003, 230). The fifth type of use is persuasive or symbolic use (Greene, 1988; Weiss

1998a) when an evaluator is hired to evaluate a program to legitimize a decision that has already been made prior to the commissioning of the evaluation. Flyvbjerg et al. (2012, 284-5) argue that in contrast to top-down utilization of applied research and evaluation knowledge the emphasis should be more in bottom-up contextual and action-oriented knowledge by asking and answering value-rational questions that stand at the core of phronetic social science[9].

This chapter has discussed the factors that enable and limit utilization of evaluation findings. The problem with most utilization-related debates is that they consider evaluation an intrinsic system that is immune to external influences. In real-life situations, this is not true. The limits and boundaries of evaluations depend on the nature of problems or policies (simple vs. wicked) and of the operating environment (static vs. turbulent). Therefore, evaluators need to tailor their approach to fit the complexity of the circumstances they face (Kaivo-oja & Stenvall 2013).

Addressing this need, Snowden and Boone (2007) broadened the traditional approach to leadership and decision-making by drawing upon complexity science to develop the Cynefin[10] framework, which allows decision-makers and leaders to see things from new viewpoints, assimilate complex concepts, and address real-world problems and opportunities. According to Snowden and Boone (2007, 69), "simple and complicated contexts assume an ordered universe, where cause-and-effect relationships are perceptible, and right answers can be determined based on the facts. Complex and chaotic contexts are unordered—there is no immediately apparent relationship between cause and effect, and the way forward is determined based on emerging patterns. The ordered world is the world of fact-based management; the unordered world represents pattern-based management." When adjusting decision-making strategies to changing environmental conditions, it is important to make a distinction between objective changes and changes in decision-makers' perceptions. Sense-making has been defined as "the ongoing retrospective development of plausible images that rationalize what people are doing" (Weick et al. 2005, 409). The concept was introduced to organizational studies by Karl E. Weick in the 1970s and he aimed to encourage a shift away from the traditional focus of organization theorists on decision-making and towards the processes that constitute the meaning of the decisions that are enacted in behavior. Table 2 describes the main elements of the framework, lists the logic of sense-making and action and finally the type evaluation plan or design suitable for each category.

9 Phronetic social science is an approach to the study of social phenomena based on a contemporary interpretation of the classical Greek concept phronesis, variously translated as practical judgment, practical wisdom, common sense, or prudence. Phronesis concerns values and interests and goes beyond analytical, scientific knowledge (episteme) and technical knowledge or know how (techne) and it involves what has been called "the art of judgement," that is to say decisions made in the manner of a virtuoso social actor Flyvbjerg (2016).

10 *Cynefin*, pronounced ku-nev-in, is a Welsh word that refers to the multiple factors at play in our environment and our experience that influence us in ways we cannot understand.

Table 2. Cynefin Framework and alternative evaluation designs (Adopted and modified from Snowden and Boone 2007; Patton 2011).

KNOWN – SIMPLE **Characteristics:** - Cause and effect relationships are linear, predictable and repeatable - Policy contents are clear and easy to define - Things are known and best practices can be applied **Management or policy strategy:** - Process management and re-engineering; incremental policymaking Sense --> Categorize --> Respond **Evaluation Design:** - Summative evaluation - Validation of best practices - Focus on monitoring, outputs and outcomes	**KNOWABLE – COMPLICATED** **Characteristics:** - Cause and effect are separated over time and space - Policies are interconnected and difficult to separate - Scenario planning, foresight and systems thinking needed **Management or policy strategy:** - Systems management; experimental policymaking Sense --> Analyze --> Respond **Evaluation Design:** - Formative evaluation, realist evaluation - Focus on context, mechanisms - Apply systems thinking
UNKNOWABLE – COMPLEX **Characteristics:** - Cause and effect is contingent and are only coherent in retrospect and non-repeatable - Policies are blurred and problems are wicked **Management or policy strategy:** - Complex adaptive systems and pattern management; agile Policymaking - Probe --> Sense --> Respond **Evaluation Design:** - Developmental evaluation - Relational approach and complexity thinking - Focus on interconnections, patterns and change paths - Identification of the leverage points in a system	**UNKNOWABLE – CHAOS** **Characteristics:** - No cause and effect relationships are perceivable. - Rapidly changing and highly unstable systems dynamics - Focus on interventions **Management or policy strategy:** - Crisis management, intuition and emergency policymaking Act --> Sense --> Respond **Evaluation Design:** - Rapid assessments - Situation awareness, validity of the data available - Focus on utilization, speed and clarity

Table 2 describes the main characteristics of the four cells of the Cynefin Framework. In the original framework there is also a fifth domain - disorder. That context makes it particularly difficult to recognize when one is in it. Action patterns are often random and situational awareness is partly missing. Simple contexts can be characterized by stability and clear cause and effect relationships that are easily discernible by everyone. According to Snowden & Boone (2007) simple contexts require straightforward management and monitoring. This is possible because cause and effect relationships are linear, operating environment is relatively predictable and repeatable and policy policy contents are clear and easy to define. In this field of "known knowns," decisions are easy to make because most parties share a common understanding. The role of public managers or policymakers is to sense, categorize, and respond. Management and policymaking is normally incremental and follows the principles of process management. According to Patton (2011, 109) the most suitable type of evaluation in simple context is summative evaluation. The role of evaluator is to monitor implementation and focus on efficiency, quality and best practices.

In a complicated context, however, decision-making and evaluation are more difficult due to the presence of "known unknowns." A complicated context is the domain of expertise, in which decision-makers can identify appropriate solutions by analyzing, investigating, and diagnosing. Causes and effects are separated by time and space, and policies are interconnected and difficult to separate. Therefore, scenario planning, foresight, and systems thinking are crucial. The role of the policymaker is to sense, analyze, and respond. This approach is not easy and often requires expertise, either in-house or external. To support policymakers, evaluation needs to be more formative, take context and mechanisms into account, and follow realist evaluation principles. Also, systems thinking is needed to understand the interconnections between policies and the interaction patterns of agents participating in the policymaking process.

In the complex context of the Cynefin framework, the whole is much more than the sum of its parts. Complex contexts are characterized by emergence and feature no best practices or right answers, but many "unknown unknowns." Much contemporary policymaking has shifted towards these contexts. Most situations and decisions in organizations are complex because major changes cause unpredictability and flux. In a situation like this, cause and effect are normally contingent and are only coherent in retrospect and in a non-repeatable sense. Policies are difficult to frame and define, and problems are typically wicked. Therefore, policymakers need to probe first, sense second, and respond third. In other words, complex contexts call for intelligent public policy (see Virtanen and Stenvall 2018). Developmental evaluation (see Patton 2011) seems to be most suitable for complex policymaking situations. In developmental evaluation evaluators should apply a relational approach and complexity thinking and focus on interconnections, patterns, and change paths. Another important role for evaluators in a complex context (e.g., an organization, policy domain, economy, or ecosystem) is to find leverage points in the system at which a small shift in one factor can produce widespread changes.

Finally, in a chaotic context, it is impossible to find right answers or solutions. In this context of "unknowables," no cause and effect relationships can be perceived due to the rapidly changing and highly unstable system dynamics. Policymakers must concentrate on quick fixes and political rescue operations. The evaluator's role is to carry out rapid assessments and try to clarify situational awareness.

The next chapter examines how systemic thinking and complexity can lead to additional interpretations of the problem of causal explanation described above and to better utilization (and understanding) of the information provided by an evaluation. It presents the basics of system thinking and complexity theory and their application to policy and program evaluations. Finally, examples of how the systemic evaluation model can be applied in practice are provided. The case study involves thematic evaluation of the disaster preparedness and risk-reduction activities of the Philippines in 2017.

4 SYSTEMS APPROACH AND COMPLEXITY

4.1 SYSTEMS THINKING AND EMERGENCE

Systems thinking and complex adaptive systems research or complexity science is a loosely organized academic field that has grown up around the study of complex systems. It is not a single theory but encompasses more than one theoretical framework and is highly interdisciplinary, seeking the answers to some fundamental questions about living, adaptable, changeable systems (Stroh 2015).

According to Ståhle (2004, 3), complexity thinking has changed the system paradigm in three ways. An understanding of system equilibrium, persistence and continuity has changed to imbalance, continuous change and discontinuity, where chaos and disorder were found to be a natural state and inevitable cause for systems development. The perception of ways of influencing the system changed so that its external manipulation and objective monitoring were seen to be impossible or useless efforts and that the system could only be influenced from the inside. In order to get systemic information, we should be the interacting as part of it. Third, the task and the attention of the research were structured in a new way, instead of truths and coherent prescripts, it was meant to understand the dynamics of the organization and radical change processes.

Ben Ramalingam uses in his book "*Aid on the Edge of Chaos*" systems thinking and complexity concepts to reveal the deep reasons and underlying patterns for why development aid either works or doesn´t (Ramalingam 2013). It also provides an informative tour of the reductionist thinking and over-simplistic approaches that characterize so much current development policy and practice. Harry Jones argues in his Oxfam blog[11] that the reason complexity matters, comes down to one or more of three major challenges:

1. Distributed capacities: the knowledge and capacities required to tackle problems are spread across actors without strong, formalized institutional links.
2. Divergent goals: inherent to many problems are divergent interests, competing narratives or conflicting goals.
3. Uncertain change pathways: it is unclear how to achieve a given aim in a given context, or change processes involve significant, unpredictable forces.

11 https://oxfamblogs.org/fp2p/complexity-101-behind-the-hype-what-do-we-actually-know/

Michael Q. Patton (2011) distinguishes three key elements that characterize complex adaptive systems:

1. Nonlinearity - small actions can produce large reactions
2. Emergence - patterns emerge from self-organization among interacting agents
3. Dynamic adaptations - interacting elements and agents respond and adapt to each other

Systems thinking deals with relatively stable patterns and trajectories and history deals with particular events, variations and decisions, shocks that take place in particular places at particular times – but complexity thinking marries the two and provides us with sophisticated unique theory of change. As Boulton et al. (2015, 29) put it: "It is detail and variation coupled with interconnection that provide the fuel for innovation, evolution, change and learning". Systems thinking can be defined as a cognitive process of studying and understanding complex systems. There are several definitions available e.g.:

- Barry Richmond (1994) defines systems thinking as the art and science of making reliable inferences about behavior by developing an increasingly deep understanding of underlying structure.
- Peter Senge (2006), another leader in the field, defines systems thinking as a discipline for seeing wholes and a framework for seeing interrelationships rather than things, for seeing patterns of change rather than static snapshots.
- Linda Sweeney and John Sterman (2000) authors and researchers in the field of systems thinking, found that that much of the art of systems thinking involves the ability to represent and assess dynamic complexity (e.g., behavior that arises from the interaction of a system's agents over time). They list specific systems thinking skills as including the ability to:
    - Understand how the behavior of a system arises from the interaction of its agents over time (i.e. dynamic complexity);
    - Discover and represent feedback processes (both positive and negative) hypothesized to underlie observed patterns of system behavior;
    - Identify stock and flow relationships;
    - Recognize delays and understand their impact;
    - Identify nonlinearities;
    - Recognize and challenge the boundaries of mental (and formal) models.

Compared to linear Logical Framework, Logic Model or Results Chain approaches, the systems models focus more on transformation mechanisms that translate inputs into outputs and outcomes. Logic model thinking considers the end product to be

sum of the actions, whereas systems thinking sees that has emerged as a product of interactions.

There has been an increasing criticism that RBM (Results-Based Management model) and logic model approach in evaluation are not sufficient tools for studying or evaluating development interventions in complex settings (see e.g. Patton 2011; Ramalingam 2013). Rationalistic (in the meaning of synoptic) planning frameworks, which embed the causal logic behind actions (from inputs to outcomes and impacts) have been developed and used extensively, especially in the field of international development aid. Too often the inflexibility of the LFA can limit staff capacity to adapt to emergent trends by holding them accountable to predicted cause and effect rather than accountable for the ability to learn from the use of rigorous evidence analysis in implementation and to adapt to changing circumstances and understand the conversion mechanisms that translate inputs into outputs, outcomes and impacts (Virtanen & Uusikylä 2004).

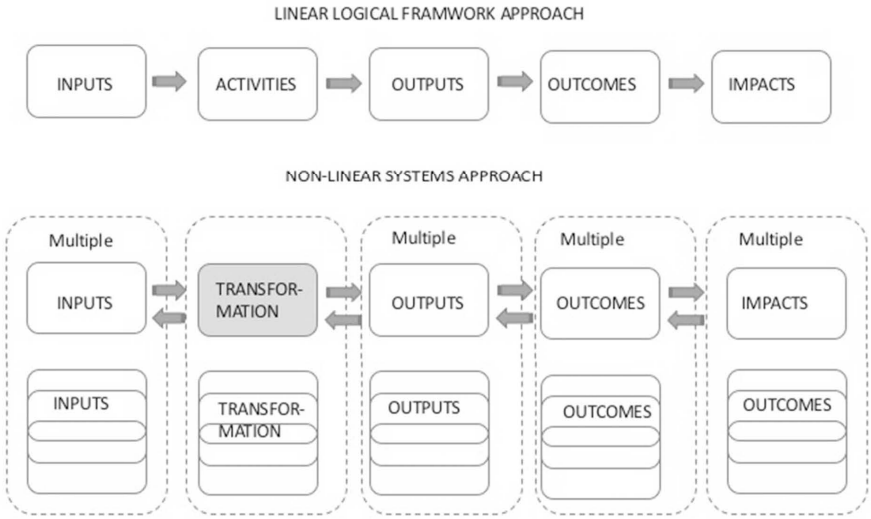

Figure 11. The intervention logic of the Logical Framework Approach (LFA) and Systems Approach (SA) (source: the author)

The Logical Framework Approach treats the process planning and implementation as linear process where A leads to B and B leads to C etc. whereas in Systems Approach A can lead to B and C simultaneously and B and C can fire back to A. Policymaking platform can be seen as a complex adaptive system in which, a group of agents (policymakers, civil servants, firms, NGOs, lobby organizations and specialized interest groups) interact in interdependent ways to produce system-wide patterns, such that those patterns then influence behaviour of the agents. We can call this nested structure policymaking culture or commonly agreed policy or strategy.

Table 3. Differences between Traditional Program Approach and Systems Approach (source: the author)

ELEMENTS	PROGRAM MODEL	SYSTEMS MODEL
Intervention Logic	Linear	Non-linear
Idea on results	Predetermined and fixed	Emerging and changing
End product	Sum of the actions	Product of interactions
Key actors	Defined stakeholders	Nodes of the network
Project manager	Controller and coordinator	Enabler
Outcomes	As defined in the project plan	Real-life changes and outcomes
Coordination mechanism	Compliance, rules	Learning, trust
Success story	Achieving results	Understanding patterns

Table 3. summarizes the differences between Traditional Program Approach and Systems thinking. Traditional program theory or Logical Framework Approach relies on linear program logic and predetermined and fixed results and outcomes that are constructed as sums of the individual actions. Actions are to be coordinated according to predefined project or program plan. According to the systems approach, social reality comprises interacting parts, which consequently cannot be first treated independently and then simply aggregated to describe the whole as in the analytical micro to macro approach. Systemic approach takes the systems approach even further by analyzing not only systems and their sub-system but also potential trajectories emerging from collision of interconnected agents in a policy space (i.e. exploration of the space of possibilities). Interconnectedness and trust are main characteristics of a complex socio-economic system. Therefore, systems analysis emphasizes concepts such as: emergence, co-evolution, connectivity, simple rules, iteration and self-organizing principles. Fryer (2013) offers a list of properties for complex adaptive systems:

1. **Emergence:** Instead of being planned or controlled, the agents in the system interact in apparently random ways. From all these interactions patterns emerge which informs the behavior of the agents within the system and the behavior of the system itself.

2. **Co-evolution:** All systems exist within their own environment and they are also part of that environment. Therefore, as their environment changes they need to change to ensure best fit. But because they are part of their environment, when they change, they change their environment, and as it has changed they need to change again, and so it goes on as a constant process. That is: communities where work takes place are not static but in a process of constant evolution and change.

3. **Connectivity:** The ways in which the agents in a system connect and relate to one another is critical to the survival of the system, because it is from these connections that the patterns are formed and the feedback disseminated. The relationships between the agents are sometimes more important than the agents themselves.

4. **Simple Rules:** Complex adaptive systems are not complicated. The emerging patterns may have a rich variety, but like a kaleidoscope the rules governing the function of the system are quite simple. From the point of view this means only some changes of perception on work.

5. **Iteration:** Small changes in the initial conditions of the system can have significant effects after they have passed through the emergence - feedback loop a few times (often referred to as the butterfly effect). This idea is likely to have effect on project methods and training.

6. **Self-Organizing:** In a pure model there is no hierarchy of command and control in a complex adaptive system. There is no planning or managing, but constant re-organizing takes place to find the best fit with the environment. In the real (organized) world this naturally is a naïve assumption but it might give some new ideas especially in developing community-based projects.
7. Edge of Chaos: Complexity theory is not the same as chaos theory, which is derived from mathematics. But chaos does have a place in complexity theory in that systems exist on a spectrum ranging from equilibrium to chaos. A system in equilibrium does not have the internal dynamics to enable it to respond to its environment and eventually slowly die. A system in chaos therefore ceases to function as a system.

8. Nested Systems: Most systems are nested within other systems and many systems are systems of smaller systems.

Williams and Hummelbrunner (2011, 18-23) suggest that the best means of understanding what is involved in thinking systematically is through three concepts or elements: interrelationships, perspectives and boundaries. Translating these elements into evaluation questions would be the following:

Interrelationships

The systems field can help in analyzing dynamic and non-linear aspects and provides a range of modelling techniques which fall into two broad categories: models that

provide insight (e.g. rich picture, causal loop diagrams, social network analysis) or models for prediction (stock-flow diagrams, agent-based modelling). The evaluation questions to be asked would be:

- How to make sense of the nature of interrelationships within a situation?
- How to make sense of the processes between them?
- How to make sense of the patterns that emerge from those processes, with what consequences, and for whom?
- Why does this matter? To whom? In which context?

Perspectives

The systems field has developed a methodology for looking at situations from different perspectives (soft systems). And it offers techniques for conveying ideas between different stakeholders (e.g. circular dialogue, systemic questioning) and to overcome differences (e.g. dialectical methods of inquiry, solution focus).'

- What are the different ways in which situations can be understood?
- How are these different understandings going to affect the way in which people judge the success of an endeavour?
- How will people´s different understandings affect their behavior, and thus the behavior of the situation, especially when things go wrong from their perspective? With what result and significance?

Boundaries

The systems field offers a methodology for assessing the consequences of boundary choices (critical systems heuristic). In addition, some methods contain techniques for addressing specific boundary issues (e.g. viable systems model, container difference-exchange model).

- How is a situation being framed, i.e. who is drawing what kind of boundary?
- What are the practical and ethical consequences of this framing, and what do those consequences imply for action?

4.2 COMPLEXITY THEORY AND CHALLENGES TO EVALUATION

It has been noted that linear or rational evaluation methods are qualified as not suitable to provide balanced judgements for wicked problems, because they only fit the unambiguous contexts of tamed or simple problems (Termeer & DeWulf 2018; Head & Alford, 2015). These methods do not account for typical wicked problem challenges such as examination of multiple frames, interconnectedness of policies, conflicting goals, ambiguous or changing priorities, problematic attributions, uncertain solutions, constantly evolving problems and time and scale disconnects (Termeer & DeWulf 2018, 4). According to Cilliers:

> A complex system can be seen as a network of dynamic nonlinear relationships. These relationships can be equated with Derrida's notion of traces. The dynamics of the system is a result of all the interactions in the system, but since this interaction also consists of multiple simultaneous nonlinear feedback, with a constant flow of energy through it, it operates in a state far from equilibrium. This perpetual activity is in effect a form of différance. This notion is extremely useful to describe the way in which the emergent properties of the system can manifest themselves, yet be in constant transformation. (Cilliers 2011, 142-143).

Complexity theory is more a loosely-coupled group of theories and models that attempts to rationalize the behavior of large and complex systems that operate at the edge of chaos. Complex systems are chiefly concerned with the behaviors and properties of systems. A system, broadly defined, is a set of entities that, through their interactions, relationships, or dependencies, form a unified whole. Origins of complexity theory can therefore be traced back into general systems theory (Bertanlanffy 1968) and cybernetics (Wiener 1950) and today it is associated mainly with the fields of computer science, mathematics and economics and biology etc. The first research institute focused on complex systems was the Santa Fe Institute (founded in 1984). Scientists associated with the Santa Fe Institute are especially Stuart Kauffman (Kauffman 1993, 1995, 2000) and John Holland (Holland 1995, 1998) that have worked on complex adaptive systems (CAS), as well as the work of scientists based in Europe such as Peter Allen (1997) and Brian Goodwin (Goodwin 1995).

Complexity studies do not provide a single theory, but the question is a combination of theoretical frameworks and trends across a range of disciplines. Complexity studies separate two subfields that examine two different kinds of emergence: the study of complex physical systems (CPS) and study of complex

adaptive systems (CAS). These elements, normally called as agents learn and adapt in response to interactions with other agents (Holland 2014, 8). Within the organizational studies, evaluation studies and among scholars of public policy, complexity has not only been seen an emerging theory of governance, but also as a powerful metaphorical tool. Kurt A. Richardson has explored the theoretical grounds of the complexity studies in the field of organization science. He defines three main approaches to study complex systems in the field of organizational studies: 1) The reductionist or neo-reductionist approach, 2) metaphorical approach and 3) critical pluralistic approach (Richardson & Cilliers (2001) and Richardson (2008)). The Neo-Reductionist School represents a kind of fierce complexity that believes in modeling and explaining their power. This community seeks to uncover the general principles of complex systems, likened to the fundamental field equations of physics. It seeks to identify and model the natural principles that affect the functioning of socio-economic systems. The Metaphorical School represents the soft line of complexity sciences. In this, the concepts offered by complexity thinking are utilized symbolically as stimuli stimulating thought. The purpose is not to model things but to look at them from new perspectives. According to this school, the complexity perspective, with its associated language, provides a powerful lens through which to see organizations or governance systems. The Critical Pluralistic School, which is also called complexity thinking, rejects the modeling of the complexity of the idea of a neo-ideological approach. Representatives of the school also resort to a relativism typical of the metaphorical school, where complexity concepts are used rather casually to describe almost any phenomenon. The Critical Pluralist school is not so interested in phenomena that can be explained but in phenomena that can not be explained. The school focuses on the epistemological consequences of complexity, in other words how we can gain understanding of complex reality and what can we all know about.

During last decays complexity science has been gaining traction also in the field of organizational studies in terms of self organizations and management (Byrne 1998; Stacey & Mowles 2016) and more generally in social sciences (Innes & Booher 2010; Luhman 1995; Boulton et al. 2015; Watts & Strogatz 1998). By studying complicated, complex and chaotic systems, scholars of social sciences can overcome an atomistic view of the rational choice thinking where society is often seen as a playground occupied by atomistic individuals, their (self)interests and resources available. Complexity theory encourages us see how order, pattern and structure arise or emerges from set of interactions. We can see how the many disparate elements of a system work with each other to shape the whole and its outcomes, as well as how each element evolves over time.

As Paul Cilliers (1998, 107) defines the analysis of complex systems:

> "In our analysis of complex systems [...] we must avoid the trap of trying to find master keys. Because of the mechanisms by which complex systems structure themselves, single principles provide inadequate descriptions. We should rather be sensitive to complex and self-organizing interactions and appreciate the play of patterns that perpetually transforms the system itself as well as the environment in which it operates"

Daviter (2017) has put forward three different ways to face wicked problems. These are problem-solving, taming, and coping. Basically, no strategy is better than the other. Instead of choosing a single correct strategy, the question is rather about the compromises between the different strategies and whether different responses to differently puzzling problems are needed. Discussion about complexity in relation to evidence-informed decision-making has been emphasized in recent years (Newman & Head 2017, Ansell & Geyer 2017, Wolf 2018). Evidence-based decision-making clearly shows two clear schools. On one had rationalists believe in the gathering of objective evidence and emphasize the primacy of the fundamentality of decision-making. Constructivists (based constructivist knowledge theory) on the other hand see evidence as being subjectively subjected to subjective interpretation, emphasizing democratic decision-making and warning of the risks of top-down rationalist decision-making (Raisio et al. 2018b; Newman 2017).

The value of the complexity theory is that it helps to seek a balance between the two schools and rise above the confrontation. Raisio et al. (2018) interpret that rationalistic and constructivist views are not as incompatible as are generally assumed (see also Newman 2017). More likely complexity thinking helps to identify different types of problems and to position rationalist and constructivist views in relation to different decision-making situations. Raisio et al. (2019)have elaborated the dichotomy between tame and wicked problems, sketching the so-called wickedness cube, which also recognizes messy problems. In the model, the tame, messy and wicked problems are perceived in relation to three dimensions: 1) the uncertainty in relation to risks, consequences of action and changing patterns; 2) complexity of elements and interdependencies; and 3) divergence and fragmentation in viewpoints, values and strategic intentions. These dimensions form the wickedness cube in which the simple and tame problems lie in the lower left corner, messy and complicated problems near the center of the cube and wicked and complex problems in the top right corner of the cube. It should be noted that this is an ideal model. In reality, the difference between different types of problems is not as clear as it appears in the cube. The boundaries of the types of problems can be blurred and can change as part of the development of the situation or phenomenon. It is also

possible to develop the presented theoretical model, for example, by adding other types of problems (Head & Alford 2017) or dimensions such as time to the cube. For the empirical applications of the wickedness cube, see Raisio et al. (2018).

Identifying wicked problems dates back to the late 1960s (Churchman 1967). Ritter and Webber (1973, 160) conceptually distinguish two extreme categories of problems in decision-making situations (i.e., tame problems and wicked problems). The main characteristics of wicked problems according to Ritter and Webber (1973) are listed in Raisio et al. (2019, 5):

1) There is no definitive formulation of a wicked problem. Different approaches to the problem see it differently. Different proposed solutions reflect the fact that it is defined differently.
2) There is "no stopping rule." Unlike in an experiment where you can stop natural processes and control variables, you cannot step outside a wicked problem or stop it to contemplate an approach to answering it. Things keep changing as policymakers are trying to formulate their answers.
3) Solutions are not true or false, rather they are good or bad. There is no right answer, and no one is in the position to say what is a right answer. The many stakeholders focus on whether proposed solutions are ones they like from their point of view.
4) There is no test of whether a solution will work or has worked. After a solution is tried, the complex and unpredictable ramifications of the intervention will change the context in such a way that the problem is now different.
5) Every solution is a one-shot operation. There can be no gradual learning by trial and error because each intervention changes the problem in an irreversible way.
6) There is no comprehensive list of possible solutions.
7) Each wicked problem is unique, so that it is hard to learn from previous problems because they were different in a significant way.
8) A wicked problem is itself a symptom of other problems. Incremental solutions run the risk of not really addressing the underlying problem.
9) There is a choice about how to see the problem, but how we see the problem determines which type of solutions we will try and apply.
10) Wicked societal problems have effects on real people, so one cannot conduct experiments to see what works without having tangible effects on people's lives.

The critical element in the analysis of complex systms is feedback. According to Richardson (2008, 14) "it is the existence of nonlinear feedback in complex systems that allows for emergence, self-organization, adaptation, learning and many other key concepts that have become synonymous with complexity thinking—and all the things that make management such a challenge. It is not just the existence of feedback loops that leads to complex behavior". That leads us to policy-learning and

the role of evaluation supporting learning and development in complex adaptive systems.

4.3 EVALUATING NETWORKS AND COMPLEX ADAPTIVE SYSTEMS

Systems thinking and complexity theories have recently also been applied in the field of policy analysis and evaluation (Patton 2011; Boulton et al. 2015; Reynolds 2015; Eoyang & Oakden 2016). This has created a wide range of options for evaluating complex interactions and emergent outcomes in program implementations. In complexity evaluation (Stame 2004; Byrne 2013; Patton 2011), the evaluator cannot turn to the traditional program of reductionism and its techniques of controlled experiments and retrospective statistical evidence. Byrne (2011) states that researchers (or evaluators) should focus on the trajectories of systems understood as descriptions through time of the nature of the systems and should track the points of stability and change in the systems.

Byrne and Callaghan (2014) argue that, for restricted complexity, the specification of rules driving system behavior can be understood as equivalent to a description of causality.

1) For complex systems, what is caused is the state of the system, the character the system has at any given time point in its trajectory.
2) There is no simple direction of causality in any sense for complex systems. Intersected systems have causal powers in relation to one another. Cause operates in any and all directions.
3) Causes in relation to complex systems are seldom if ever single or additive. Interaction among elements is always what matters.
4) For complex systems of the same kind. The same system state may be produced in different ways, and those ways are not only multiple in number but generally are complex and interactive in kind.
5) Time matters. Sequence and duration both have to be considered when exploring causality in complex systems.
6) The point of hunting causes is only secondary to explain what is in terms of retroductive history. The interesting thing is the informing of action directed toward the achievement of futures.

Innes and Booher (2010) state that when applying complexity theory to public policy and planning, one should focus on the interactions and relationships rather than the system as a whole. They summarize five key features of CASs that need to be considered when studying or evaluating complex policy domains and phenomena.

Table 4. Main evaluation areas of complex adaptive policy systems (Modified from Innes & Booher (2010, 32).

Feature	Description	Evaluation criteria
Agents	The system comprises large number of individual agents connected through multiple networks.	Inclusiveness of the networks. Dynamics of the systems. Number and type of agents.
Interactions	The agents interact dynamically, exchanging information and energy based upon heuristics that organize the interactions locally. Effects propagate though the system.	Type of interaction patterns and networks (weak ties or strong ties; loosely coupled or dense networks). Dissemination mechanisms and patterns. Role of subgroups or individual agents (brokerage, isolated nodes, centrality etc.)
Nonlinearity	The interactions are nonlinear, iterative, recursive, and self-referential. There are many direct and indirect feedback loops.	Type of dynamics Feedback mechanisms
System behaviour	The system is open, the behaviour of the system is determined by interactions, not the components, and behaviour of the system can only be understood by looking at the interactions.	Interaction patterns Learning type (single loop, douple loop or triple loop learning)
Robustness and adaptation	The system displays both capacity to maintain its viability and the capacity to evolve. When the agents will adapt to each other, the system can reorganize its internal structure without the intervention of an external agent.	Capacity to adopt and evolve Level of trust Number of enabling or restricting structures or processes.

In contrast to the traditional approaches, developmental evaluation by Michael Quinn Patton applies complexity concepts and approaches to enhance innovation and use of evaluation findings (Patton 2011). Context is a very crucial term also for Michael Quinn Patton, widely considered as the founder of the current Developmental Evaluation (DE) paradigm. In his 2011 textbook he specifies some ideas behind his thinking. Table 3. describes the main evaluation areas of complex adaptive systems. The main elements of the table are actors, interconnections, nonlinear dynamics, system behavior based on interactions and adaptations mechanisms. According to Patton (2011, 1) developmental evaluation supports innovation development to guide adaptation to emergent and dynamic realities in complex environments. Innovations can take the form of new projects, programs, products, organizational changes, policy reforms, and system interventions. A complex system is characterized by a large number of interacting and interdependent elements in which there is no central control. Patterns of change emerge from rapid, real time interactions that generate learning, evolution, and development – if one is paying attention and knows how to observe and capture the important and emergent patterns. Complex environments for social interventions and innovations are those in which what to do to solve

problems is uncertain and key stakeholders are in conflict about how to proceed. Hayes (2008, 417) argues that evaluation of policy processes based on complexity theory emphasizes the need for a whole systems analysis that can be informed by a time series perspective. Only this transitional and analysis approach (see chapter 5.) will give the full understanding on complexity and its impact on public policies.

Developmental Evaluation

Developmental evaluation emphases innovation and strategic learning rather than the standard outcomes and logic model-based approaches discussed earlier in this paper. In this sense, it resembles the so-called realistic evaluation model. Pawson and Tilley developed the first realistic evaluation approach in 1997. They argued that in order to be useful to decision-makers, evaluations need to identify what works in which circumstances and for whom, rather than merely asking "does it work?" The complete realist question is: "What works, for whom, in what respects, to what extent, in what contexts, and how?" In order to answer that question, realist evaluators aim to identify the underlying generative mechanisms that explain how the outcomes were caused and the influence of context (Pawson and Tilley 1997).

Emergent and dynamic realities in complex environments and interacting and interdependent elements are fundamentally new thoughts in the program evaluation tradition, which normally follows the causal explanatory path and focuses mainly on deviations of the pre-set targets or benchmarks. Although Patton is not explicit in explaining the mechanisms that transform or give form for the new emerging elements, he hints that these result from a complex set of interaction between actors, ideas and competing preferences or issues.

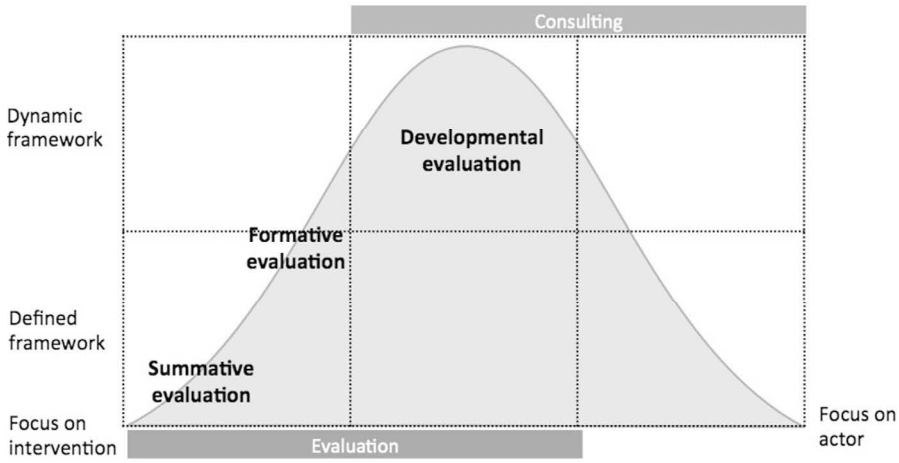

Figure 12. Developmental evaluation as dynamic reframing (Adapted from Patton 2015, copyright Front Ltd).

Figure 12 illustrates how developmental evaluation is especially needed when actions are embedded in a complex system, and actors (policymakers, experts, program managers, etc.) are trying to change it. Actors make these decisions with limited information and knowledge, which most often leads to unintended consequences. Therefore, decision-making is constant learning (usually by trial and error), and that is why developmental evaluation is a process of dynamic reframing. When choosing the "right" evaluation approach, the level of complexity and the role of the evaluator should be taken into consideration.

Traditional program evaluation approaches can be categorized roughly into two different approaches: formative evaluations and summative evaluations. Formative evaluations (or *ex ante* evaluations) are typically conducted before a program is launched in full-scale. The purpose of a formative evaluation is to improve the program model. Summative evaluations (or *ex post* evaluations) are conducted after the program (or some phases of it) has ended. The purpose of a summative evaluation is to assess whether the program has been successful. Developmental evaluation differs from both these approaches as it aims to develop the whole process (goals, methods, etc.) continuously to best respond to the changing conditions.

The main differences between developmental evaluation and traditional program evaluation are summarized below:

- Formative evaluation aims to improve and fine-tune program
- Summative evaluation tests, proves and validates program models
- Linear problem solving and known cause of the problem, high predictability
- Programs are outcome-driven, how best to reach the defined goals
- Developmental evaluation aims to continuously develop the whole process; both the goals and the methods, to best respond to the changing conditions
- Nonlinear, complex and dynamic conditions for problem solving, high unpredictability
- Programs are driven by the aim to enable systems-change
- Social innovations and adaptive management
- Developmental evaluation can also lead to generation of a model to be evaluated formatively and summatively.

Figure 13 suggests that another criterion for choosing the most appropriate evaluation design is the phase or life stage of an initiative or program. Developmental evaluation is most suitable for new, explorative and innovative programs that experiment with different approaches and activities. In these circumstances, developmental evaluation can report the results and experiences of the pilots and, thus, give reflective support to program management to redesign the program targets.

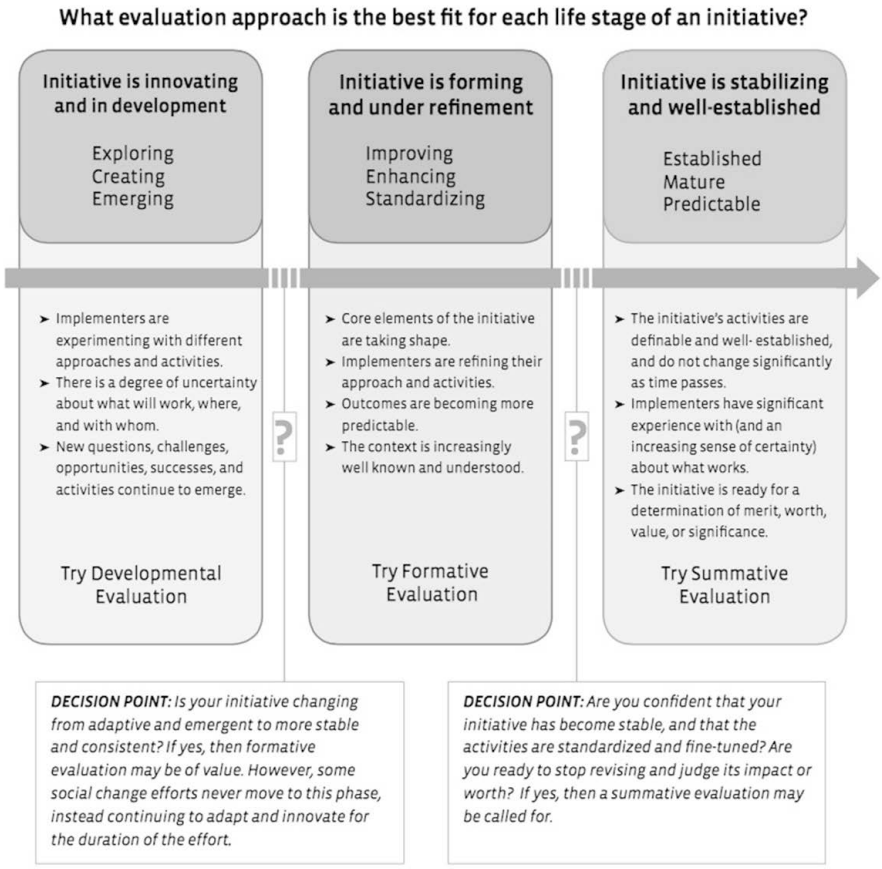

Figure 13. When to choose developmental evaluation approach (Preskil & Beer 2012, 6)

Above, developmental evaluation as it relates to summative and formative evaluation approaches was explained. Yet, there is still confusion as to how other complexity-driven, systems-oriented or real-time evaluation models relate to developmental evaluation and what the major differences between these are.

Outcome Mapping and Outcome Harvesting

One approach to make a conversion mechanism explicit is to apply outcome mapping (OM). The objective of the OM method, as with contribution analysis (CA), is to improve causal interpretation in the simple program model and logical framework. Outcome mapping also attempts to highlight nonlinear effects and individual functional and behavior-related changes in the groups targeted by the measures.

Williams and Hummelbrunner (2009, 77) raise four key principles of the OM method:

1) *Actor-oriented development and changes in behavior*. Outcome mapping takes into account the fact that people and organizations manage change processes. A program's underlying problems and goals and the related indicators must be determined by means of the changes observed in actors' behavior.
2) *Continuous learning and flexibility*. The model emphasizes that the most functional and effective planning, monitoring and evaluation measures are always cyclical, iterative and reflexive. They attempt to increase knowledge of the actors, contexts and challenges involved with the push for social change.
3) *Participation and accountability*. Having partners and interest groups participate in the planning and monitoring process and emphasizing the reflexivity and reciprocity of interdependency creates shared commitment and accountability for the implementation of the program.
4) *Nonlinearity and contribution*. The starting points for change processes and transformation are shared collectively. Instead of funder-controlled, mechanical input–output thinking, these are the result of complex interdependencies.

Outcome Harvesting (OH)has proven to be useful in complex environment or with complex policies when it is not possible to define concretely most of what an intervention aims to achieve, or even, what specific actions will be taken over a multi-year period. Methods enables evaluators, program managers and other stakeholders (also participants) to identify, formulate, verify, and make sense of outcomes. The method was inspired by the definition of outcome as a change in the behaviour, relationships, actions, activities, policies, or practices of an individual, group, community, organisation, or institution. Unlike some evaluation methods, Outcome Harvesting does not measure progress towards predetermined outcomes or objectives, but rather collects evidence of what has been achieved. (Wilson-Grau 2015.). For applications of Outcome Harvesting in complex evaluation settings see Uusikylä et al. (2017).

Network analysis

Social network analysis (SNA) is a method that has been applied in sociology, organizational research and policy studies for a long time to analyze complex interactions[12]. Network analysis is a cross-disciplinary endeavour to understand

12 To learn more about the background of social network analysis and its applications and methods see e.g. Johanson et al. 1995; Wasserman & Faust 1994; Degenne & Forsé 1999.

how social systems work and what is their impact on communities and societies as a whole.). It has become increasingly common to apply the program evaluation framework (e.g. Caloffi 2015; Uusikylä et al. 2017; Durland et al. 2005). Social network analysis is based on an assumption of the importance of relationships among interacting units. The social network perspective encompasses theories, models, and applications that are expressed in terms of relational concepts or processes. Along with growing interest and increased use of network analysis has become a consensus about the central principles underlying the network perspective:

- Actors and their actions are viewed as interdependent rather than independent, autonomous units
- Relational ties (linkages) between actors are channels for transfer or "flow" of resources (either material or nonmaterial)
- Network models focusing on individuals view the network structural environment as providing opportunities for or constraints on individual action
- Network models conceptualize structure (social, economic, political, and so forth) as lasting patterns of relations among actors

The unit of analysis in network analysis is not the individual, but an entity consisting of a collection of individuals and the linkages among them. Network methods focus on dyads (two actors and their ties), triads (three actors and their ties), or larger systems (subgroups of individuals, or entire networks) (Wasserman & Faust 1994). In the following section, the systematic evaluation methods such as outcome harvesting or network analysis are applied in the evaluation of complex interventions (disaster preparedness and risk reduction in the Philippines). The purpose of this chapter is to demonstrate that the systematic evaluation models need not just remain at theoretical or conceptual level but can be tested in real life situations.

4.4 CASE STUDY: SYSTEMS EVALUATION ON DISASTER PREPAREDNESS

This case study demonstrates through systems lenses, how an evaluation can be undertaken in complexity. In the section that follows, a case study provides an overview of how systems evaluation approach was used:

- To assess local level dynamic disaster processes and interactions between formal and informal disaster preparedness actors.
- To identify and interpret multiple and interconnect risk-patterns and
- To evaluate emergent process at the community level

In 2016 group of evaluators lead by Petri Uusikylä assisted the Finnish Red Cross (FRC) to develop a systemic evaluation approach to identify and address critical emerging from the introduction of disaster preparedness and risk-reduction programs implemented by the FRC. Uusikylä et al. (2017) introduced an alternative perspective to study disaster preparedness and disaster risk reduction (DP/DRR) systems. A thematic Study shows that by applying systems thinking and complexity theory we understand better the dynamics and interconnectedness of the DP/DRR. This applies both to interconnected risks (multi-risk landscapes) and Interconnected actors (multi-actor networks). These results were part of the broader study commissioned by The Finnish Red Cross (FRC).

The aim of the thematic study was to promote institutional learning on successful/unsuccessful Disaster Preparedness and Disaster Risk Reduction (DP/DRR) project experiences and practices that can benefit better programing in the future. The thematic study consists of two main parts. First part presented the results of the meta-analysis of the ten countries and 17 projects. Meta-analysis utilizes the IFRC/MFA evaluation criteria (relevance, impact, effectiveness, efficiency, sustainability and coherence). From this sample the final case studies were selected. Last part is the case study section introduced the findings and results of the field missions to the Philippines, El Salvador, Honduras and Tajikistan[13]. Case study analyses use set of systems methods and tools to better understand the dynamics and interconnection between the risk factors and stakeholders in the field. The systems approach utilized in case study section gave interesting insights of the dynamics and inter-connectedness of risk landscapes and inter-organizational DM networks. Study showed that by applying systems methods such as network analysis the risk components helped local disaster risk management units to better understand the interconnectedness of risk elements and the joint impact on those risks. Also, the relations and connections between the disaster risk agencies and stakeholders better explain why certain risk preparedness actions produce better results and effects. The more actors are connected to the network the more versatile the understanding of the risk preparedness and thus higher the resilience of preparedness actions.

The evaluation team carried a field mission to Aklan Province and Caloocan City in the Philippines on 14-19 November 2016. During the field mission The Philippines Red Cross (PRC) staff (HQ, Chapters and volunteers) were interviewed together with main stakeholders and beneficiaries (52 in total). Assessment of the main risks related to both communities was one of the interview topics. Instead of repeating the risk part of the Vulnerability and Capacity Assessment (VCA), a more comprehensive systems approach to risk identification and analysis is introduced.

13 Similar case studies were carried out also in El Salvador, Honduras and Tajikistan. See full report in Uusikylä et al. (2017) or analysis based on systems approach in Uusikylä et al. (2020).

This approach was warmly welcomed by the participants of the Risk Assessment Workshops in Caloocan City and Kalibo (Aklan).

Multi-Risk Assessment

Evaluation showed that VCA process helps to identify the main risks that are more or less known a priori but most severe damages are normally caused by simultaneous, interlinked risks. Risk definition should therefore be broadened to cover also political, economic, health-related, social, technological and cultural risks. Secondly, risks should not be seen as independent or atomistic phenomena but tightly interlinked web of risk patterns. The risk assessment framework of the World Economic Forum (see e.g. WEF 2014) was applied. In addition to natural disaster, also the so called systemic risks that have sometimes explanatory power when explaining human behaviour in critical disaster environments were included. Systemic risk is the risk of "breakdowns in an entire system, as opposed to breakdowns in individual parts and components" (Kaufman and Scott 2003). Systemic risks are characterized by (WEF 2014):

- modest tipping points combining indirectly to produce large failures
- risk-sharing or contagion, as one loss triggers a chain of others
- "hysteresis", or systems being unable to recover equilibrium after a shock

According to the Community-based Disaster Risk Reduction Project 2015 Annual Report the planning process in Caloocan City remained to be difficult. Only one community was able to complete their Barangay DRRM 2015 plan, however, copy could still not be secured from the community or city. And only 2 communities (12 and 176) were able to finalise their contingency plan (these were also the Barangays that participated in workshops).

Evaluation found that the most critical risks are: 1) floods, 2) earthquakes and 3) fire. Given the urban context these are also the most dangerous for the inhabitants and livelihoods in the area. However, when mitigation measures are considered it is not sufficient to tackle these risk factors separately. Most of the risks listed above are highly interconnected and therefore also mitigation measures, drills and evacuation plans should be planned from the multi-risk perspective.

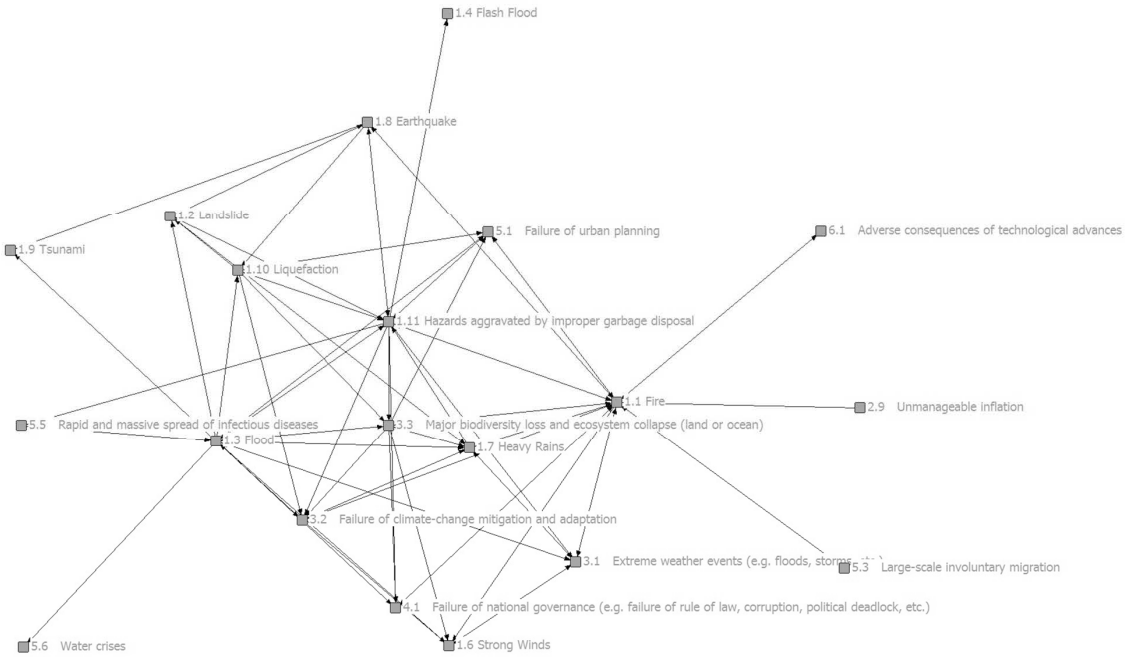

Figure 14. Caloocan multi-risk map of interconnectedness (Uusikylä et al. 2017).

In Figure 14 the numerous and complex interconnections between risks can create consequences that are disproportionate and difficult to contain or predict. The Risks Interconnections Map (RIM) seeks to connect the dots by identifying and visualizing the underlying patterns. This allows for a better understanding of the impact of systemic risks so as to mitigate them by identifying the transmission channels between risks and potential second- and third-order effects. These interconnections do not represent direct causality. They are likely to be indirect, for example through parallel impacts or mitigation trade-offs. The RIM map shows how all global risks are connected to others and underlines the complexity of dealing with global risk in an effective manner. The map visualizes the strength of connection between individual risks – the most strongly connected risks could merit additional attention due to the multiple ways they affect or are affected by other risks.

Interestingly, hazards aggravated by improper garbage disposal (listed as No. 4 in prominence) together fire (No. 2) and major biodiversity loss ecosystem collapse (No. 10) seem to be the most centrally positioned risk factors that bridge several other risk areas. These risks together with failure of the national governance are not typically listed in VCAs or other risk assessment exercises. During the focus groups discussions many participants announced that when they carry out risk assessments next time they are willing to apply a broader, systemic multi-risk approach. This

also indicates that there should be special risk identification maps and mitigation measures for Urban DP/DRR projects.

The systemic multi-risk mapping widened the risk landscape both in Aklan and in Caloocan City. It showed that risk such as failure or urban planning and infrastructure are directly connected phenomena such as hazards aggravated by improper garbage disposal and liquefaction. Likewise, failure of national governance directly increases major biodiversity loss and failures of climate change mitigations. When planning future urban DRR project these second layer risk factor should be embedded into overall risk landscape and taken into consideration when planning project activities.

Social Network Analysis

In the complex world, relations (collaborative and competitive) between individuals and organisations need to be given a special attention. The introduction of general systems theory into discussions of management theory served as a stimulus, and a way forward, for those seeking to explore relations between organisations. They described systems theory as a way of integrating diverse internal and external factors that faced managers. In their view, systems theory also helped managers to cope with the complex nature of these factors. There are two dimensions across which organisations can be related. They can have: interactive relationships, for instance in the exchange of information or resources or non-interactive relationships when they share particular attributes—such as status, identity, cognitive structures, strategic positioning, or core technology—that induce the same behavioural stimuli in related members and/or expose the organisations to the same evolutionary forces (see e.g. Gharajedaghi 2011, Holland 2014). Here the focus is on on direct interactions between the organisations and group of actors in Caloocan and Aklan DP/DRR networks.

During the workshops and interviews, the participants were first asked to list the most important/prominent organisations and groups in their local DP/DRR networks. Participants were given the list of organisations and they could add additional organisations to this list. This produced a traditional stakeholder mapping. Stakeholders were asked to list and rate the most influential DP/DRR groups and organisations in their area. According to these ratings the most prominent actors in Caloocan list consists of government organisations or service agencies (Barangay LGUs, Schools, DRRM Office and Division Office) PRC Chapter being the rank 5 followed again by Bureau of Fire Protection and DILG). In Aklan, the two most prominent organisations or groups are PRC Chapter and Barangay residents and after LGUs again PRC 143 members and volunteers. In addition, the

Europe Aid project has reported the strong institutional flavour of the Caloocan DP/DRR culture.

After listing all the relevant stakeholder related to DP/DRR activities in Caloocan City and Aklan, the respondents were given a matrix (with additional stakeholders included) and asked to define the relationships between all actors (organisations and groups) in the matrix. Individual Excel matrices were thereafter aggregated (mean) to form a synthesis matrix containing all the responses. This matrix was then copied to UCINET program for network analysis and mapping. Figure 16 shows the overall structure of the inter-organisational DP/DRR networks in Caloocan City.

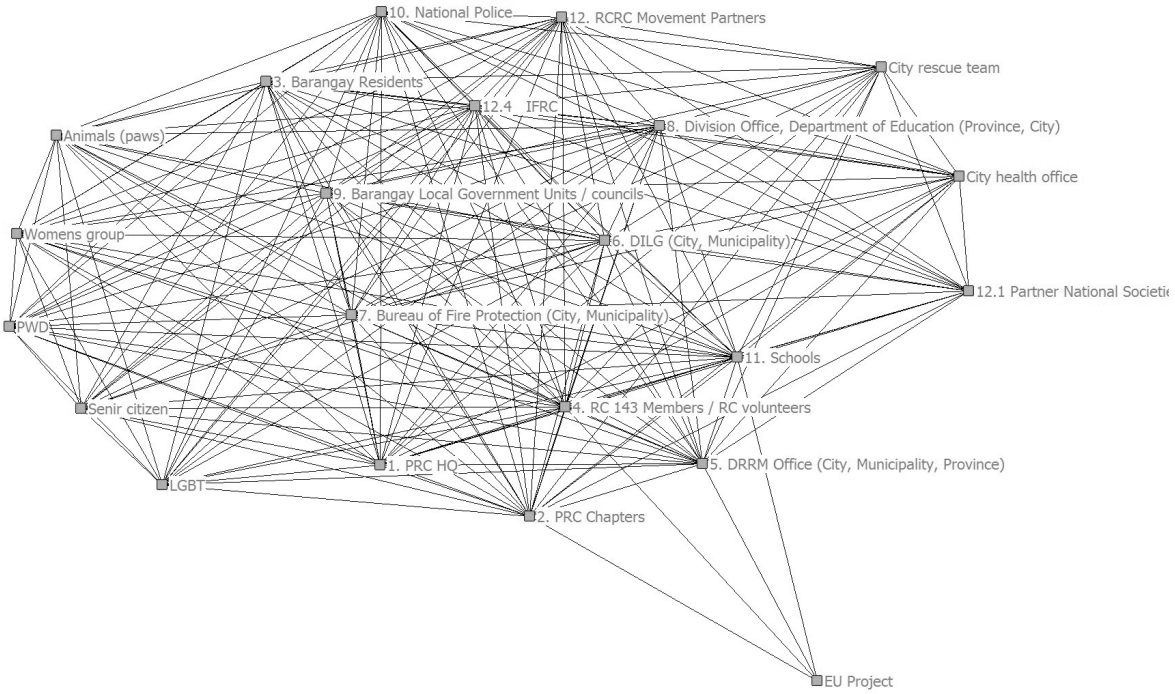

Figure 15. Inter-organizational DP/DRR Network in The Caloocan City (Uusikylä et al. 2017).

The network above (Figure 15) is very dense and highly interconnected. In practice, this means that all organisation and groups can reach each other's at least through paths (i.e. via brokerage). Strong government driven culture in Caloocan can be seen by analysing the organisations that occupy central positions in the middle of the networks. This brokerage position increases their power in resource and information sharing. EuropeAid project states in its 2015 Annual Report that "close and formal links with the DILG and LGUs have not been established, risking that they are not very much interested in the products of the project. The project sees the LGUs as target for the envisaged advocacy activities, instead of considering them

as partners or alleys working for the same cause. The opportunities for influencing local planning have become very limited. In addition, during the field mission it could be clearly seen that Barangays that received funding from the projects were not willing to disseminate the knowledge gained to other Barangays (winner takes it all). This is not a very good cultural mode from the sustainability perspective.

Network analysis of the Caloocan city stakeholders showed that there were many groups or agencies that were rather actively participating in the implementation of the project without having formal status in the CBDRR project framework. In Aklan network (reference group) the role of other NGOs (e.g. World Vision), charities and associations (Rotary Club) and especially the Catholic Church and companies that were considered to be important actors in DP/DRR work. In Caloocan the role of the citizen groups (women groups, senior citizens) and health care agencies was more important than expected in project plans. These groups provide additional resource that gives the CBDRR project more opportunities and are likely to increase resilience in the area.

The roles of different collective actor groups in the DM networks will be discussed next[14]. Although, we have argued that DM Networks are self-organising social systems which manifests certain characteristic that may retain even if all its individual members are replaced (see Lazlo 1972), each actor occupies a certain structural position in a network that either constraint or enables actor to pursue its goals and ambitious. Consider the dynamics of the diffusion of new ideas or information in a network. Central individuals, organisations or groups embedded in a system of strong ties not only have a high potential for transmitting ideas, but also can send messages to those who share those ideas and practices (Kadushin 2012, 145). To analyse the network positions of various collective DM actors we use two indicators to measure their influence. First one is Freeman's centrality index, which measures the number of direct ties that an actor has to other members of the network. The score we have calculated from the case study DM networks is the betweennes centrality[15]. A practical interpretation for the betweenness centrality is the position of a brokerage. An actor with high betweennes centrality links groups that might otherwise not be connected. The less constrained broker or organisation that bridges structural holes (Burt 2005) can be very effective opinion leader. This gives four DM roles in a disaster preparedness and risk reduction network:

14 Network data was aggregated by coding each actor into a collective actor group. The adjacency matrix was partitioned into submatrices by computing the average scores for each subgroup. This data was thereafter used as N x N network matrix.

15 Degree centrality can be defined as the number of links incident upon a node (i.e., the number of ties that a node has). Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes. It was introduced as a measure for quantifying the control of a human on the communication between other humans in a social network by Linton Freeman. See more Freeman 1979 or Johanson et. al 1995).

- **Power Brokers** are DM actors that have *high brokerage* position but a *low centrality*. These actor normally build bridges between agents in a network. Their position could also be characterized as an enabler, that is an agent that provides possibilities, opportunities and contacts to other members of the network but does not actively play a leading role.
- **Support Agents** are DM Network members that have both *low centrality* and *low betweennes* scores. These agents typically have certain restricted or limited responsibilities or very specified tasks in a network.
- **Guardians** are well connected (*high centrality*) but do not occupy strategically central positions connecting other members of the network (*low betweennes*). In DM Networks this means that network members are not dependent on Guardians but still frequently interact with them
- **Game Changers** are the most influential actors in a DM network. Game changers have vast amount contacts with other network member (*high centrality*) and occupy strategically critical positions that gives them an opportunity channel information and other resources (*high betweennes*). Game changers are agents that can change an existing situation or activity in a significant way. On the contrary they can also block a change they are not willing to support.

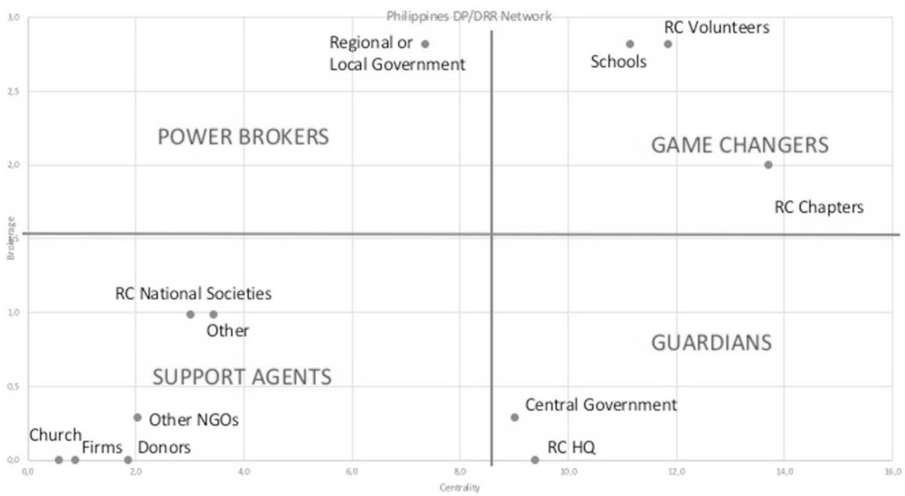

Figure 16. Collective Actors´ Roles in the Philippines DM Network (Uusikylä et al. 2017).

Figure 16 describes the collective actors´ roles in the Philippines disaster managemetn network. The Philippines DM Network is run by the RC Chapters, network of RC 143 volunteers and schools that play an important role in implementation of the DP/DRR projects and activities. Quantitative network analysis formalizes and support the qualitative information and observation gathered during the Caloocan and

Aklan field mission. The capacity of Philippines Red Cross is extremely high and its project are well prepared and implemented. This is one of the main reasons and explanation for the strong position of the RC actors in the DM Network. The role of the Philippines Red Cross Head Quarter is more of a guardian (together with the Central Government). HQ mobilizes the resources need for DP/DRR activities, is in close contact with Central Government (advocacy) and steers the activities from the arm´s length. Regional and local government actors are typical power broker. They enable DP/DRR activities and connect various actors in the field. Support Agents group in the Philippines consists of organisations such RC National Societies, NGOs, Firms, Church and other donor organisations. All these agents have a limited scope for their actions and carry out specific tasks in the DM Network.

Outcome Harvesting

One of the problems with LFA-driven approaches is that it steers focus too narrowly on the programme/project outputs and outcomes and easily neglects other important changes, results and outcomes. Outcome Harvesting is a method that enables evaluators, grant makers, and managers to identify, formulate, verify, and make sense of outcomes. The method was inspired by the definition of outcome as a change in the behaviour, relationships, actions, activities, policies, or practices of an individual, group, community, organisation, or institution. Unlike some evaluation methods, Outcome Harvesting does not measure progress towards predetermined outcomes or objectives, but rather collects evidence of what has been achieved, and works backward to determine whether and how the project or intervention contributed to the change.

Outcome Harvesting method was tested both in Caloocan City and in Aklan to measure all possible changes (either positive or negative) observed by the mission informants. Participants of the workshops were given three outcome forms each and were instructed to list the most important changes (from the DP/DRR perspectives) taken place in their operating environments. Results of the Aklan Outcome Harvesting are presented in Table 3.

Table 5. Aklan Province Outcome Harvesting Results (Uusikylä et al. 2017).

What?	Why?	What was the impact?	Who contributed? Who were the change agents?		
Behavioural change - people are dependent on outside support	Project cycle and exit	Negative			
Community resilience and preparedness	Typhoon Yolanda showed that people were more prepared	Positive	Community volunteer	Barangay officials	Barangay leaders
Low sustainability of the activities in the community	Less monitoring from the project implementer	Negative			
Participation of the community people	Active participation to activities	Positive	MDRRMO	RC143	
Good governance and awareness of the LGU and the community	Active collaboration between the PRC and LGU	Positive	Civil Society Organisations		
Development of DRRM system in Aklan	Active collaboration between the LGU and stakeholders	Positive	Private sector	International NGOs	Volunteers
DRR Awareness	When typhoon hits people know what to do	Positive	PDRRMO	MDRRMO	BDRRMO
Advocate participatory approach in planning	By disseminating information to communities by drills and simulation	Positive	LGU	International NGOs	
Public awareness has risen dramatically	When typhoon Frank hit Aklan people started to notice the importance of DRR work	Positive			
Proactive culture	The school children learned to value "preparedness through training"	Positive			
Community has become more proactive, prepared and resilient	After several trainings	Positive	MDRRMO	PDRRMO	

Table 5 reports the most relevant changes and outcomes by the respondents. All except two (behavioural change – dependency on outside project support and low sustainability) are positive. Most outcomes relate to behavioural or cultural changes such as: DRR or public awareness has risen, participation has increased, orientation and culture has become more proactive and governance and DRRM system has developed. Respondents were also asked what the main causes for these changes were. External shocks and disasters such as Yolanda & Haiyan in 2013 have been among the most important causes. Also participation to various drill and training has played a major role. Outcome Harvesting questionnaire had also one question (scale 1-10) related to the impact of externally funded projects (mainly PRC/FRC project in Aklan) on outcomes listed earlier. Figure 8 summarizes the main results of the Aklan Study.

Project(s) seemed to have rather high impact on most of the outcomes (especially on public and DRR awareness whereas proactive culture and community resilience are more multidimensional phenomena that are not only results of the successful project. Also, two negative outcomes seemed to be caused mainly by other factors beyond the project scope.

Conclusions

The evaluation found that systems thinking and complexity theory can provide useful tools for disaster preparedness and reduction. Looking the DP/DPR ecosystem as a whole and a result of interactions and interconnections helps in anticipating risks that otherwise would have gone without noticing and understanding linkages that are useful in increasing resilience in the communities. The paper finds that proactive and participatory project design and planning are key factors of successful project implementation and exit as it increases local ownership in projects and thus is likely to improve the sustainability of a project. This is also in line with the assumptions of process oriented resilience which emphasises local knowledge and culture as the basis of resilient disaster risk reduction strategies. Also, the paper states that projects contributing to long-term collaboration practices generate good results. Close collaboration and needs assessment as early as possible was emphasised continuously in the field study interviews and stakeholder workshops.

Identification of major stakeholders is very important part of successful project planning and implementation because each actor occupies a certain structural position in a network that either constrain or enables actor to pursue their goals and ambitions. In most of the case study projects the main stakeholders were identified but there were some difficulties in defining the network boundaries. The study found that inter-linkages between stakeholders were unknown and there was very little attention paid to cross-sectoral coordination with other critical policy domains (e.g. health, social sector, housing and urban planning). During the field visits, network analysis was introduced as a tool to map inter-organisation stakeholder networks, which illustrates better the complex inter-linkages of stakeholders in the ecosystem. Participants in various workshops found this a very useful and were willing to get more information and training on network analysis.

The Evaluation analysed the stakeholder networks of the case studies and report the main network structures and aggregated the data to see which organisations are acting as power brokers, support agents, guardians and game changers. In all of the case studies the RC organisations (either headquarters and/or chapters, districts or branches had a position of a game changer, which was a very positive result. Identifying stakeholder networks and relationships increases the resilience of the ecosystem as it eases collaboration and co-operation among stakeholders in a complex system.

5 SYNTHESIS: ADAPTIVE EVALUATION FOR TRANSFORMATIVE GOVERNANCE

The aim of the introductory part of this dissertation is to summarize the central message of the scientific articles behind this work in a broader theoretical context. These arguments can be stated as follows. Individual evaluations and fragmented administrative structures cannot provide a sufficient understanding of the knowledge base required for democratic decision-making; therefore, they are also unable to provide sufficient evidence of the impacts of policy decisions made. The reductionist program evaluation model has proven inadequate to assess complex policy problems and decisions. In addition, linear monitoring and evaluation models and the positivistic evaluation paradigm do not work in a rapidly changing, complex, and interdependent world. There is certainly a need for a new way of perceiving societal governance, adaptive programming tools, and flexible evaluation designs. This requires governance arrangements that are both adaptive to changing conditions and effective in dealing with multiple challenges. This chapter aims to outline the principles of transformative governance, adaptive programming, and a flexible and information-sharing evaluation approach.

Figure 17. A Synthesis: Critical elements for coping with complexity (source: the author)

Figure 17 summarizes the key findings of this dissertation to respond to the challenges of complexity from the perspective of evidence-based policymaking, evaluation, and policy learning. The first factor is the conceptualization of the systemic and transformative governance framework. If the governance model reflects a reductionist way of perceiving social change and the roles of institutions, then the subsequent developments of program activities and evaluation systems are of little importance. After the governance system has been changed into one that supports system change and policy learning, one can move on to perceive program-level change factors. As Virtanen and Kaivo-oja (2015, 82) argue "public services cannot rely on simple "best practices" or "good practices," as, in complex systems, "next practices" and "unique practices" will be much more relevant. Systemic adaptability requires new kinds of leadership and management styles, kind of new systemic governance reaching beyond the current NPG and system leadership paradigms". Adaptive programming is an interesting approach that has been applied in recent implementations of innovation projects and evaluations (see, e.g., Derbyshire & Donovan 2016; Wallenius et al. 2019). Adaptive programming means that program content is constantly reviewed based on feedback received during program execution. Thus, the role of the evaluators also changes from positivist assessors to the direction of an independent external expert who develops and facilitates. In the following subsections, these three components will be assessed in greater detail.

5.1 TRANSFORMATIVE GOVERNANCE

Walton (2016, 84) has argued that program framing and governance arrangements can provide the mechanisms for applying complexity theory within evaluation, program, and policy processes. According to Geels' (2002) multi-level perspective (MLP), systemic change is often about the interaction among global, national, and local socio-technical developments. The top level of the model is formed by the broader context (landscape) of society and societal development, which can only be affected to a very limited extent by different actors in society. The middle level (the so-called regime level) reflects the prevailing socio-political system, which is made up of interacting elements of society (e.g., politics, administration, science, and technology). The third level (niche consists of innovative experiments taking place outside the current system, which, together with the changes to the system of internal opportunities and changes to the operating environment, can lead to a wider social transformation. Transitions are thus processes of structural change in the prevailing systems (and their subsystems). Under appropriate conditions, structures can change rapidly, but often a wider social transition can take decades (Geels 2002; Geels & Schot 2007). Research has conducted learning-oriented experiments as

key elements of systemic change. Radically new operating models often originate from the integration of such experiments into other ongoing development processes (Geels 2004; Elzen et al. 2004). However, the MLP has also been considered to be too vague and abstract, neglecting the role of institutions and path dependencies and lacking an analysis of agency and economic variables (see Meadowcroft 2011; Berkhout et al. 2004; Smith et al. 2005).

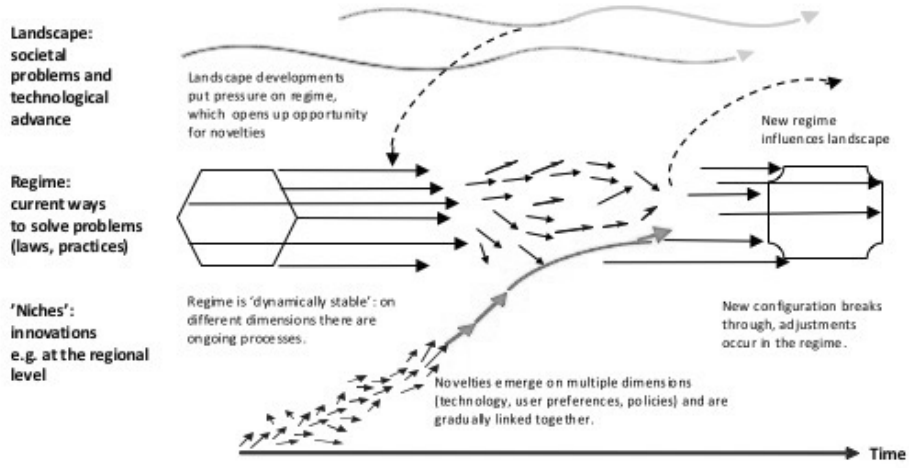

Figure 18. Multi-level perspective on to socio-technical change and complex governance (Geels & Schot 2007, 401).

Figure 18 describes the transition management idea in a broader multi-level perspective on complex governance. Transitions are transformation processes in which society changes in a fundamental way over a generation or more (Rotmans et al. 2001, 15). Transition management (TM) offers a comprehensive approach to transformative governance. TM aims to facilitate and accelerate sustainability transitions through a participatory process of visioning, learning, and launching experiments pilots. TM is based on the idea that big systemic changes cannot be controlled completely, but it is possible to control their speed and direction. TM emphasizes the importance of a shared vision and guidance, while another school using the multi-level model of social change, strategic niche management (SNM), relies more on bottom-up innovations (Loorbach 2009; Loorbach & Rotmans 2006).

According to TM's lessons, social change can be controlled through interventions at different levels. At the strategic level, it is about building consensus on the problems and trends of the operating environment. Through this, one can create a commitment and a common will for the direction in which change is to be exported. According to TM's teachings, this is done in practice by creating a common vision and starting point for a dialogue change between various stakeholders (the "previous

followers") in the "transition arena." The key is that the supreme political leadership gives support to the work of the arena of change, but without much intervention. At the tactical level, the aim of the transition arena is to translate into different networks, organizations, and institutions and to draw up a "transition agenda" and, on that basis, "Transition Paths" to achieve goals. Transition experiments at the operational level mobilize field actors extensively to implement change and transform vision objectives into concrete practical steps.

Transformative change requires addressing governance failures by integrating coordination improvements during the construction of transformative change pathways (Schot & Steinmueller 2018). Transformative governance is urgently needed to support policy learning and sustainable transition. Transformative governance is a nested, multi-level, hierarchical coordination structure in which a constantly fluctuating, interactive network of actors plays an important role. This means that it needs to be coordinated through simultaneous horizontal transactions instead of top-down hierarchical steering. This emphasizes the need for carefully balanced policy mixes for creative destruction, which is particularly important in the field of sustainability transitions (Kivimaa & Kern 2016). The second important factor is the capacity to monitor, anticipate, and involve all actors in the self-governance process of transformative change (Weber & Rohracher 2012). What is described above is how governance structures need to become more supportive of transition and change. Here, transformativity and self-organization are central topics. Next, we shall look at what this means at the level of individual policy instruments, such as programs. Adaptive program theory is a promising approach to perceive programming in a way that takes greater account of complexity, change, and policy learning.

5.2 MISSION-ORIENTED POLICIES AND ADAPTIVE PROGRAMMING

The second element of the complexity-driven transformative governance model in this context is the idea of mission-oriented policies and adaptive programming. According to Mazzucato (2018, 229-230 public service agencies have been considered as necessary but unproductive actors throughout the history of economic thought – rather than value creators. This line of argumentation leads new definition of public value and empasizes the role of public sector in achieving valuable societal goals (Mazzucato 2013). Mission-oriented policies can be defined as systemic public policies that draw on frontier knowledge to attain specific goals – in other words big science deployed to meet big problems. Missions aim to provide solutions or opportunities and an approach to address the numerous challenges that people face

in their daily lives. These challenges cannot be reduced to externalities or public goods (Mazzucato, 2018a; 2018b).

In his excellent work, "Policymaking for a Good Society" (2010), Gregory Hayden looks at the significance of a holistic approach with regard to programme evaluation. He goes deeper into societal problems and the programmes drafted to solve them by means of *metapolicymaking* and the so-called Social Fabric Matrix method (SFM). The SFM model emphasises the social institutions, cultural and individual values, attitudes and beliefs, technologies and ecological systems that affect programmes in the background. Insofar as these elements are concerned, this approach is closely related to the so-called new institutional organisational theory, even though he makes no direct references to it. Hayden also thinks that the field of societal decisions and programmes should be considered holistically; that is, taking into account all a given society's activities affecting the phenomenon in question.

Adaptive management and programming go along with the idea of mission-driven policymaking. In fact the adaptive programs can be seen as mission projects that are needed to implement mission-driven policies. Adaptive programming principles draw heavily on complexity theory and systems thinking, and they attempt to understand how interconnections between different elements affect each other and how they produce change (Bain et al. 2016, 23). In contrast to the conventional approach, adaptive programming is about designing and implementing programs through learning by doing, continually exploring, experimenting and testing, and adapting program approaches throughout delivery, allowing local partners and stakeholders to identify problems and work together to develop solutions (Derbyshire & Donovan 2016, 29). Adaptive programming emphasizes continuous learning, experimental culture, and systemic transition that is often complex, unpredictable, and rapidly changing (Ramalingam 2013; Frej & Ramalingam 2011). Kleinfeld (2015, 23) discusses the limitations of the traditional linear development policy paradigm by stating:

> "…everything about complex systems seems to make any process of design and measurement impossible. Change is nonlinear; it moves in non-incremental ways and tips and metastasizes quickly based on nothing out of the ordinary other than one more grain of sand dropping on too large a pile. Idiosyncratic, small variables matter. (…) Each situation is path dependent and even a tactic that works during one's first of programming might begin to fail by the third year, as opponents wise up and change their approach". (Kleinfeld 2015, 23)

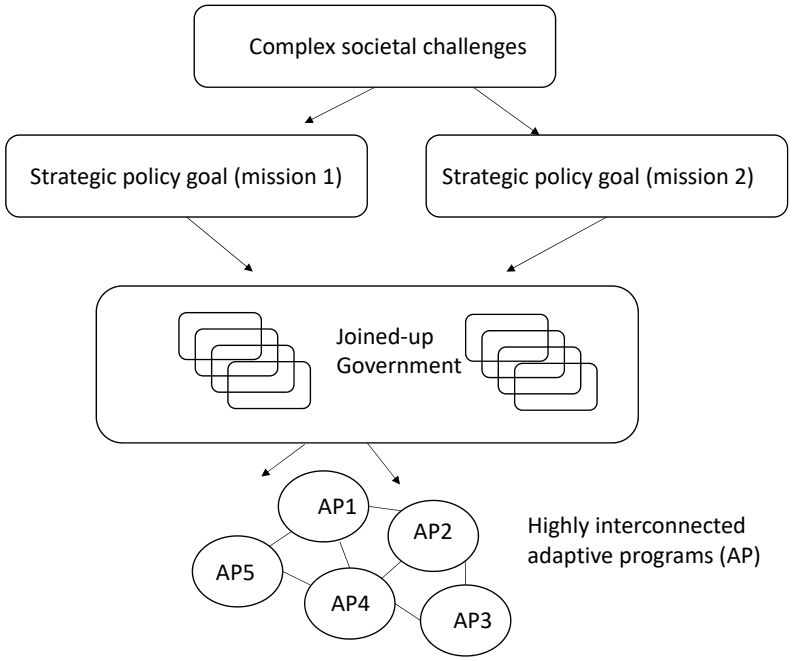

Figure 19. Mission-oriented policies and adaptive programming model (Adopted from: Miedzinski et al. 2019, 4).

Figure 19 highlights the overall architecture of mission-oriented policies and adaptive programmin. Mission-oriented policies and adaptive programming is an approach that takes contextual factors and local needs into account while emphasizing the overall performance and sustainable impacts. Adaptive management is at the heart of the new development management approach, called doing development differently. It is strongly based on complex systems analysis, participatory methods, and problem-driven and iterative adaptation (Andrews et al. 2015). The success of adaptive programming depends on how much it is mainstreamed into existing development programming, especially planning, monitoring, evaluation, and learning cycles. Oxfam's PMEL involves three strategies: (1) planning for flexibility, (2) developing locally owned monitoring and evaluation, and (3) creating an enabling environment for learning (Desai et al. 2018). Planning for flexibility refers to such strategies as experimenting with evolutionary approaches, negotiating flexible funding, and designing adaptive logical frameworks and contracts. Developing locally owned monitoring and evaluation is crucial for enhancing the participatory approach, local commitment, and ownership and action-based decision-making. Lastly, enabling the environment for leaning can be achieved by improving dialogue and communication, mentoring, and sharing mindsets, as well as enhancing the reflective project and program culture.

Adaptive programming also supports systemic program thinking and is often a necessary way of operating in a complex environment. In systems theory, the program is constantly being considered a learning and changing entity. Instead of linear, goal-bound evaluation, the systemic approach aims to understand the transformation of the program and the factors that have led to it. In systemic evaluation, through transformation and learning analysis, we can ask whether the changes made promote the relevance, effectiveness, efficiency, and impacts of the program. Alternatively, we can ask whether the changes taken moved the program further away from the goals set. For this judgment, we must rely on complex counterfactual judgments about whether and to what extent a different program choice would produce different outcomes or impacts.

Adaptive programming can be used to check program objectives and actions throughout the program cycle. At its best, this will lead to results that are even better than the original program intended. When assessing the changes or transformation to the program logic, it is important to ask whether the changes made were proactive (something was done in advance to produce better results) or reactive (something was done because external changes already took place). Self-organization is the basic characteristic of complex systems and adaptive programming. It is the ability of systems to develop in time and place without external intervention through internal fluctuations. Two characteristic features are: 1) the core of the interaction. The basic essence of the system cannot be understood if it does not investigate the interaction that occurs in it. 2) A general view of the system cannot be obtained from outside. The process of the emergence of relationship patterns is a key element of self-organization (Ståhle 2004). According to Boulton et al., it is called self-organization because:

> "the patterns of relationships that emerge are a) not designed by an external agency, b) what form they take cannot be predicted, and c) they do not in general accord with any overarching principle such as maintaining stability or maximizing profits or minimizing energy" (Boulton et al. 2015, 17).

Self-organization and emergence, however, also involve problems of which one should be aware. These mainly relate to the concept of accountability. If the objectives and measures of the program are constantly changed during the program, this might lead to a situation in which the "new program" no longer resembles the original program, i.e., the content that was behind the funding decision. In this case, the accountability of the program and the administration may be jeopardized. It may sometimes be impossible for an outsider to know whether changes during the program were due to changes in real needs or whether they were the result of reactive drifting due to external pressures and signals. Program changes are

therefore always a trade-off between innovativeness or adaptivity and accountability or efficiency. Figure 20 below describes the trade-off between these two elements.

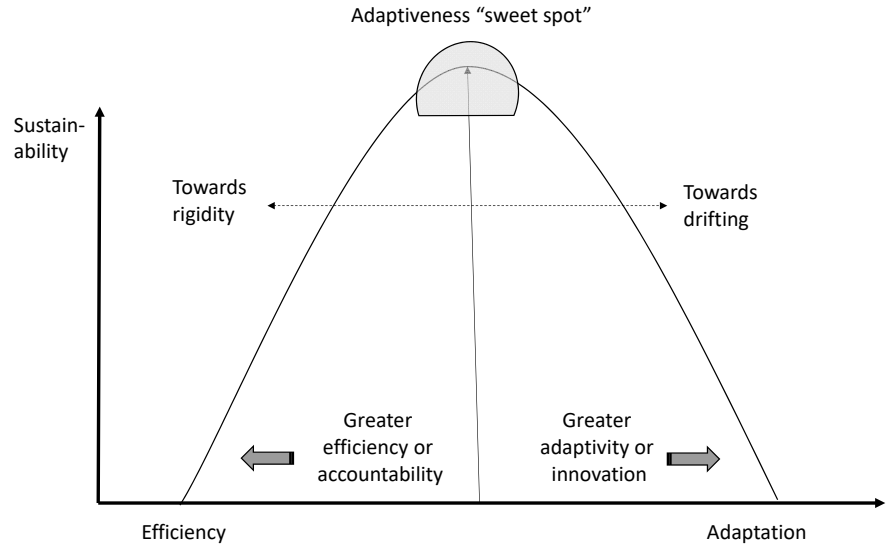

Figure 20. Trade-off in adaptive programming: accountability vs. innovativeness (source: the author)

Figure 20 shows a typical trade-off between accountability (or efficiency) and innovation (or adaptability). Ideally, the program objectives and the nature of the activity and the degree of complexity of the operating environment determine the balance between accountability and innovation. The general assumption is that the more familiar the content area of the program (e.g., technical infrastructure projects), the more static the operating environment, the more predictable the nature of the activity, and the more the focus will shift to efficiency and accountability. This also supports findings by Jalonen (2007) according to which interaction processes have unknown potential which can be translated into resources balancing effectiveness with creativity. Correspondingly, in new, complex programs, it is likely wiser to emphasize innovation, adaptability, and learning. Wallenius et al. (2019) found that in the final evaluation of the Vietnam Innovation Partnership Programme between Finland and Vietnam, the positive effects achieved by the program were largely based on the adaptive implementation of the program, which was justified in the case of a new creative (generating a start-up culture) activity.

5.3 EVALUATORS AS KNOWLEDGE BROKERS

It has been argued that to understand complexity and evaluate complex intervention evaluators, there is a need to think more strategically and apply systemic concepts and tools. Strategic thinking requires an understanding of the dynamics and interlinkages of the operating environment (e.g., the policy arena or ecosystem) and the nature and substance of an organization's knowledge assets, as well as deep engagement with systems thinking.

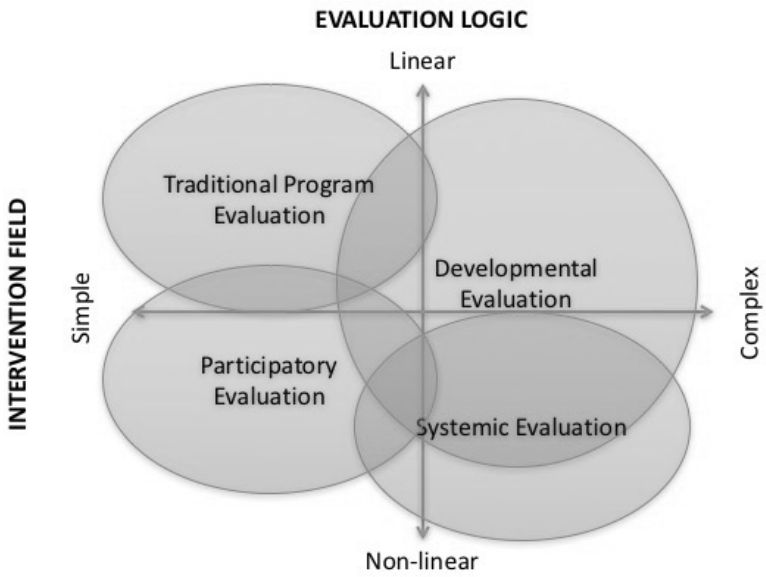

Figure 21. The nature of problems or program design vs. evaluation logic (Uusikylä et al. 2016)

To cope with the complexity, both program and evaluation design must be creative and experimental. This dissertation has mainly discussed the complex interventions and the importance of applying reflective evaluation approaches, such as developmental evaluation, to support program learning. Complex evaluation differs from simple intervention evaluation by providing real-time feedback and helping the program management team and steering group to redesign the program and make strategic adjustments when needed. Traditional program evaluation and even participatory evaluation are too slow and rigid of tools for providing real-time evidence and feedback. Figure 21 describes how different evaluation approaches best suit the four fields based on their degree of complexity and causal logic. The traditional, positivistic evaluation paradigm has conventionally been applied to situations in which policy problems are tame and the operating environment

is simple. The participatory or constructivist approach works best in a simple environment with an emphasis on building consensus and interpreting the meaning and intentions of the intervention. Developmental and systemic evaluation models are needed when the problems behind policies or program intervention are more complicated and complex. There, the developmental and systemic approaches differ mainly in their theories of change, i.e., whether the intervention logic is considered linear or non-linear. Adaptive evaluation is a relatively new concept. Eoyang and Oakden (2016) have found it useful to identify patterns in complex programs and to understand the dynamics in ways that are meaningful to stakeholders. In their work, "synergy has emerged between complexity theory (through the lens of human systems dynamics) and evaluation practice (through a case study of a complex program of social change)" (ibid, 2.).

What is the role and importance of evaluation in the future? What kind of information is needed in the future to support decision-making? Will artificial intelligence (AI) and machine learning replace evaluation in the future? For these questions, there are no unambiguous answers. AI researchers (e.g. Tuomi 2018) have stated that AI is certainly a substitute for current simple and technical expert tasks. In the area of evaluation, monitoring and simple, goal-oriented assessment could be something left to AI. However, more complex assessment methods and policy impact assessment models that require deep learning cannot be replaced by simple algorithms.

The role of the evaluator is not always clear in a multi-layered and complex policy process. Evaluators often get involved before and after the actual evaluation (Forss & Schwartz 2011, 7). The role of developmental evaluation is to track emergent and changing realities and feedback meaningful findings in real time so that reality check facilitates and supports the dynamics of innovation. Naturally, this also transforms the role and position of the evaluator from independent expert or judge to facilitator and learning coach who is part of the development team. Nutley (2003, 12) argues that the interaction between policymakers and researchers is limited by the divergence of these two worlds. They use different languages, have different priorities, different agendas, different time scales and different reward systems. Consequentially a communication gap exists.

Figure 22 aims to simplify the alternative structural positions the evaluator shares between the citizen and decision-maker. This is to demonstrate how these simplistic strategic alliances in three planning models deepen the information gaps between citizens and the government.

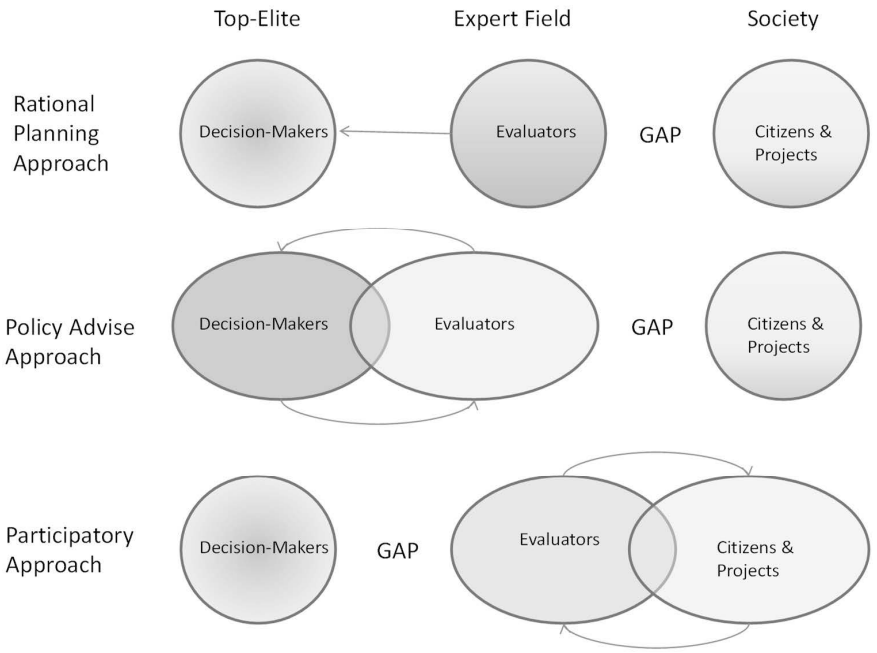

Figure 22. Evaluators taking part in strategic alliances (Uusikylä & Virtanen 2008)

The first setting, namely the Rational Planning Approach (RPA), has its roots in the positivistic evaluation paradigm. Here, the role of evaluation is mainly to judge and value policy decisions (ex ante, interim, or ex-post). Evaluators are expected to be neutral and base their scientific inquiry on pure facts and evidence. It is a well-known caveat of RPA that no matter how systematic and methodologically pure the scientific inquiry is, it is impossible to observe the reality based on hard facts only. Full-scale analysis and judgments are always dependent on the interpretative framework and subjective assessment made by the evaluator. When these are not explicitly stated, it is difficult for public audiences and citizens to verify whether evaluations are methodologically just and sufficiently robust. This creates an information gap between decision-makers (top-elite) and the civil society; thus, evaluation does not sufficiently serve the purpose of enhancing mutual learning or accountability.

During the 1980s and 1990s, rigid positivistic models were mostly replaced by constructivist (Guba & Lincoln 1989), realistic (Pawson & Tilly 1997), or utilization-focused approaches (Patton 1997). The last one emphasizes strongly the usability of evaluation results, i.e., results could be used to reformulate policies or re-structure organizations. This links the role of the evaluator to our second model, which is the Policy Advice Approach (PAA). In this approach, decision-makers (top-elite) and evaluators often belong to same policy communities and issue networks, and they are likely to share common knowledge and policy framing and thus have mutual

interests. This is likely to reduce the critical role of the evaluator and tie him/her tightly to the existing advocacy coalition trying to enhance the efficiency and effectiveness of the policies. Joint action in any closed circle of the policy community tends to create a homophily bias in networks, which means that relations are more likely between people who share socially or politically significant attributes, such as ideological values, profession, age, educational background, or place of living. Dense networks and commonly shared norms and values tend to weaken the innovativeness of the social system (Granovetter 1973; 1985) and reduce transparency and open dialogue at the systemic level. Citizens or grassroots-level projects are seldom involved or consulted. This is mainly due to a lack of information and knowledge, but is also a natural cause of complexity and multi-layer institutional settings. This is likely to widen the gap between policymakers and citizens; thus increasing passiveness and distrust in the government among citizens.

The Participatory Approach (PA) refers to a variety of methodological choices to involve and empower participants and stakeholders toward project implementation. These models also attempt to build trust relations between the evaluator and citizens or local projects. The evaluator is likely to assume a responsive, interactive, and orchestrating role, bringing together different groups of stakeholders with divergent views for mutual exploration and to generate consensus. The evaluator plays a key role in prioritizing the views expressed and 'negotiating' between stakeholders. The problem however can be that the influence of this local-level consensus building or bonding on policies normally remains rather modest. The vertical gap between implementing agencies (local governments, projects, associations) and decision-making bodies still exists despite good will to enhance the openness and transparency of the policymaking environment. In addition, the flock of evaluators seems divided into two sub-clusters: those advising policymakers at the top of the decision-making hierarchy and those empowering implementing agencies at the grassroots level. The participatory approach encourages a holistic perspective on challenges and opens up the process to a variety of innovation pathways relevant for the missions. The shared understanding of challenges and shared ownership of missions resulting from the participatory processes can improve implementation of innovation projects on the ground Miedzinski et al. (2019, 38).

As noted, there are several information or knowledge gaps between key players in society. These gaps are both horizontal and vertical. Horizontal gaps refer to a lack of *social capital*[16] among actors at the same level of governance system (i.e., European level, central government, or regional/local levels), while vertical gaps refer to a reduced amount of *policy capital* between decision-makers and civic actors.

Ronald S. Burt (2005) created a powerful theory on brokerage and closure as a means to enhance social capital. In his theory, brokerage is considered the activity of the people who live at the intersection of social worlds and who have a vision advantage of seeing and developing innovative and good ideas. Closure is the tightening of coordination in a closed network of people, and people do this well as a complement to brokers because of the trust and alignment they create.

His structural-hole argument draws on network concepts that emerged in earlier sociological writings, most notably Granovetter's (1985) writings on the strength of weak ties, Freeman's (1979) writings on betweenness centrality, Cook's (1982) writings on the benefits of exclusive exchange partners, and Burt's (1992) writings on autonomy created by complex networks. In his view:

> "a structural hole is exists between people or groups when either party is unaware of value available if they were to coordinate on some point. This leaves open the possibility that people or groups a potentially valuable context for action, brokerage is the action of coordinating across the hole with bridges between people on opposite sides of the hole, and network (…) brokers, are the people who build the bridges. (…) The social capital of structural holes comes from the opportunities that holes provide to broker the flow of information between people, and shape the projects that bring together people from opposite sides of the hole." (Burt 2005, 25-27).

By using Burt's notion of structural holes, we propose that the evaluator's role should be developed strongly toward bridging immanent structural holes (not only among actors but also between policies) that exist in modern societies. Therefore, evaluation should contribute more to strengthening a) social capital among policy-related societal actors and b) enhancing policy coherence by bridging gaps between different policies and policy arenas.

The first endeavor (strengthening social capital) would move evaluation toward deliberative policy analysis (Hajer & Wagenaar 2003; Thompson 2008; Setälä 2017)

16 There are several definitions of social capital. The World Bank defines social capital as follows: "The social capital of a society includes the institutions, relationships, attitudes and values that govern interactions among people and contribute to economic and social development. It includes the shared values and rules for social conduct expressed in personal relationships, trust and a common sense of 'civic' responsibility, that makes a society more than a collection of individuals" (World Bank 2001).

and thus aim toward strengthening democratic governance in the network society. In practical terms, it would mean shifting the focus from impact analysis toward policymaking practices, processes, and interpretation schemes. The goal is the strong inclusion of those affected by public policy and the search for an appropriate way of involving many others that might be indirectly affected by public policies This would mean the creation of well-considered linkages among citizens, policy institutions, and often-unstable policy practices (Hajer & Wagenaar 2003).

The second goal (enhancing policy capital) refers to the increasing need for a more comprehensive evaluation focus. The notion of policy capital does not have a common meaning. Therefore, it can be described as tThe overall amount of cohesion and policymaking capacity in the system. It also refers to overall knowledge and understanding on policies, degree of openness and transparency, and the system and right and opportunities to discuss and argue on policies. Evaluation should consider thematic sets of policies and their inter-linkages, not only single-policy effects and would-be impacts. In practice, this would mean that in the future, evaluations are more cross-sectoral and thematic and apply new techniques of meta-evaluation and scenario work developed in the field of futures studies. This would also mean that the applied social science-based disciplinary approach (including rigid methodology, quantitative methods, and linear causal models) should be completed by synthesizing analyses and the provision of meta-knowledge.

Evaluators as knowledge brokers (Ward et al. 2009; Meyer 2010) offer a promising new approach to the to under-socialized program approach and systems theory by emphasizing the active role of the evaluator. According to Meyer (2010, 1) knowledge brokers are people or organizations that move knowledge around and create connections between researchers and their various audiences. When applying this approach, the policy is neither a process nor a system, but consists of instantaneous and local interpersonal interaction processes, which also evolve constantly (Stacey 1996). Therefore, the policy interpretation is about dynamic self-organizing interplay patterns and meanings that are constantly being edited in speech and narratives. The model has been transformed from the theory of complex adaptive systems (CAS) to respond better to the human and social characters of life. The democratic evaluator recognizes value pluralism and seeks to represent a range of interests in his or her issue formulation. The basic value is an informed citizenry, and the evaluator acts as a broker in exchanging information between groups who want knowledge of each other (MacDonald 1978). The democratic evaluator's role is first to help all our peoples to choose between alternative societies and to provide a disinterested source of information about the origins, processes, and effects of social action. Essentially, these recommendations support a discursive or deliberative democracy (Hanberger 2004).

5.4 CONCLUSIONS

This dissertation has discussed the premises, opportunities, and constraints of evidence-based decision-making from an evaluation research perspective. In the summary and underlying articles, it has been pointed out that the shift from centralized planning and policy guidance to decentralized decision-making (e.g. New Public Management and Network Governance) has highlighted the need to demonstrate the performance and effectiveness of public policies. This change has also meant an increase in the role and importance of evaluation in all OECD countries during the last 25 years.

Since the beginning of the 2010s, the political and administrative environment has become more unpredictable and difficult to govern. The global economy, the development of information and communication technology, and social media have contributed to increasing the interdependence of societal phenomena and heightening the sensitivity of change in the operating environment. Researchers talk about the ever-growing challenge of wicked problems and complexity. How can we use evaluation information—and what kinds of evaluation information—in a rapidly changing, strongly interdependent, and complex world?

The main findings of the dissertation are as follows:

1) **Political decision-making is very much dependent on the** single issues at stake and **changes taking place in the decision-making environment**. As Hyytinen (2017) argues, "The core problem is that traditional evaluation practices do not support innovation because the systemic perspective is neglected in the targets, and the implementation of evaluation does not take into account its dynamic nature." Complex or so-called wicked problems require a different preparation process than simple and technical policy issues. Likewise, the nature of the operating environment (i.e., stable vs turbulent) has a significant impact on the dynamics of political preparation. In general, the nature of the political environment has changed from slow and predictable (cf. planning economy) to complex and unstable; in it, policy outcomes and impacts are difficult to predict (Easley & Kleinberg 2010; Byrne 2005).

This has also had an impact on the governance structures of society. The top-down bureaucratic steering model has been replaced by results management other decentralized models of democratic governance and network management. Article 3, "Transforming Silo-steering into Performance Governance System: The Case of the Finnish National Government," examines the ability of the Finnish government to utilize the monitoring and evaluation information gathered in a performance management system. The analysis shows that the evaluation data is fragmented, and its use is very limited. One of the reasons for this is the silo management and

steering structure through which each ministry or government agency looks at the impacts of policies mainly from the perspective of its own activities, programs and projects.

Article 2, "Three Spheres of Performance Governance - Spanning the Boundaries from Single-Organization Focus Toward a Partnership Network," continues this debate by highlighting how one of the reasons for unsatisfactory performance is the single-sighted and narrow scope of performance management models applied in most OECD countries. Performance targets are normally set from a single organization's perspective, which is not sufficient given the complex nature of societal problems with which public agencies deal. In the worst case, single-unit performance targeting and maximization of an organization's self-interest produces suboptimal results rather than a positive impact on society as a whole.

2) **Single evaluations tend to provide a fragmented evidence-base for policymaking.** Generating a more reliable overall picture generally requires the accumulation of cumulative data and knowledge either by synthesizing the research conducted in the policy area or by commissioning meta-analyzes and meta-evaluations that combine the results of individual evaluations and their policy recommendations. Meta-evaluations are also necessary to ensure the quality of the evaluations and the functionality of the processes. All of this is designed to increase the reliability of the underlying information and support policy learning. Article 4, "Meta-Evaluation as a Tool for Learning," argues that "To be able to utilize evaluation findings, three critical conditions have to be met. First, there should be a general understanding of the results. Second, the findings of the evaluations have to be timely, valid and reliable, and in a form in which they can be communicated to decision-makers and to the general public. Finally, findings, results and conclusions have to be legitimate (both substantively and in the processes through which they are created) and to follow sound and ethical guidelines for conducting evaluations."

Article 1. ("Exploring the Missing Links Between Causes and Effects - New Conceptual Framework for Understanding Micro-Macro Conversions in Program Evaluation") argues most of the evaluations focus on programs and their societal goals while ignoring the societal context and the actions of individuals. through which programs are implemented. According to Virtanen & Uusikylä (2004,83), "this may, in extreme cases, lead to false and mechanistic interpretations of the results and impacts of programs". Also, Pawson and Tilley (1997) suggest that an evaluator should focus on certain themes in carrying out evaluation assignments, including embeddedness, mechanisms, contexts, regularities and change.

According to Article 3, "Transforming Silo-steering into Performance Governance System: The Case of the Finnish National Government," "A critical question still is how to create systems where performance monitoring and evaluation provide both tactical and strategic feedback information on the results achieved/not achieved, and

analytical assessment behind both the failure and success. At this time government needs to support wide-ranging theoretical discussions and attempts to promote procedures for accomplishing this multi-actor assessment and interpretation." In all, the precondition of evidence-based or evidence-informed decision-making is a sufficient understanding of the complicatedness and complexity of policies and their context. This requires a phenomenon-based approach (Lähteenmäki-Smith & Virtanen 2018; Sitra 2018) to understand policy contents and interconnectedness and the interdependencies between policies, policy arenas and actors. Moreover, it is important to understand that the whole is more than the sum of its parts.

3) In complex (or even chaotic) decision-making situations or dealing with wicked problems, there is normally no clear evidence-base or quick solution available to solve the problem. Therefore, the underlying cause–effect relationships cannot be determined by mechanistic, linear models, but in order to understand complex political phenomena, the policy researcher or evaluator must seek to apply systems thinking and systemic tools to analyze or evaluate policymaking processes. In the complex contexts the whole is much more than the sum of its parts. Complex contexts are the domain of emergence, where there are no best practices or right answers. This is the field of "unknown unknowns," and it is the domain to which much of contemporary policymaking has shifted. Most situations and decisions in organizations are complex because major changes cause unpredictability and flux. In a situation like this, cause and effect is normally contingent and are only coherent in retrospect and non-repeatable sense. Policies are difficult to frame and define and problems are normally wicked. Therefore, policymakers need to probe first, then sense, and then respond. in a chaotic context, it is impossible to find right answers or solutions. In this situation of unknowables, no cause and effect relationships are perceivable due to rapidly changing and highly unstable systems dynamics. Policymakers concentrate on quick fixes and "rescue" operations. Complicated contexts are the domain of expertise, where decision-makers can identify right solutions by analyzing, investigating and diagnosing while in chaotic contexts the role of an evaluator is to carry out rapid assessments and try to clarify the situation awareness.

This is the main argument of Article 1. ("Exploring the Missing Links Between Causes and Effects - New Conceptual Framework for Understanding Micro-Macro Conversions in Program Evaluation"). The article highlights the complexity and multi-dimensional nature of modern society and argues that this is difficult to capture with mechanistic evaluation approaches. Evaluators, therefore, need specialized knowledge to calibrate their evaluation studies and design experiments. The article refers to Coleman's macro–micro conversion model which highlights the micro-macro transition as one of the main theoretical problems of social science. Coleman (1990) argues that explanations of system behavior in terms of lower-level

constituent elements are apt to be more general and more useful for interventions than those that do not probe beneath the system level. The model describes the cause and effect relationship between social structure and reciprocal behavior.

Article 2, "Three Spheres of Performance Governance – Spanning the Boundaries from Single-Organization Focus Toward a Partnership Network," further elaborates the need to replace mechanistic logic models with more system-oriented tools for planning and evaluation. The article states that it is apparent that causal logical thinking is an inseparable part of any intentional action trying to achieve set objectives. However, the authors have seriously considered the lessons coming from the abundant discussion concerning the network society, complexities of network governance and the importance of forming strategic partnerships between public and private actors in order to achieve socioeconomic objectives. The model created by the authors is called the Three Spheres of Performance Governance Model, which aims to bring together ideas from the literature of both policy studies and research on management and organization. It is an attempt to move toward greater recognition of the distinctive features of public policy management and governance.

4) **Evaluation is always affected by the political context in which it is produced and used to propagate the values of those who produce and commission it**. Therefore, the positivistic evaluation paradigm often fails in real-life situations. The absolute independence between the evaluator and the object is an ontologically unsustainable starting point. The constructivist approach (Guba & Lincoln 1989) is also associated with problems, the most important of which is the blurring of evaluation interpretation when the evaluator is too close or part of the subject. The model of developmental evaluation can be considered to be a kind of synthesis between a positivist and constructivist evaluation paradigms. Moreover, the developmental assessment has been found to be suitable for the assessment of complex phenomena (Patton 2011).

Article 1, "Exploring the Missing Links Between Causes and Effects - New Conceptual Framework for Understanding Micro-Macro Conversions in Program Evaluation," describes four alternative models (scientific inquiry, explorative inquiry, symbolic evaluation, transparent democracy) of evaluation in terms of "how explicitly causality has been taken into consideration and how the model of evaluation serves the aims of enhancing public-sector accountability. Accountability in this sense can be considered as a macro-political concept that monitors the degree of success of political and administrative attempts to provide public services to citizens and govern legitimately." Article 4, "Meta-Evaluation as a Tool for Learning," argues that meta-evaluation or meta-analysis should be used as a tool to improve the validity of evaluation findings and the utilization of the evaluation findings and recommendations.

The dissertation has shown that evaluation models and the data content they produce have not been able to develop and improve in the same way as the operating environment has changed. Monitoring and evaluations continue to be subject to overly mechanistic input-output-outcome thinking. Such linear causal thinking does not correspond to the emergent effects of the complex world or the interaction patterns between phenomena and actors. Too often, evaluation information is produced on a program basis rather than a need-and-phenomenon basis. This has led to the low utilization of evaluation data.

The dissertation emphasizes that the problem mentioned above is not only due to evaluations; rather, the entirety of our program and project-based development thinking should be reconsidered. Current program thinking is largely based on the planning ideals of the 1960s and 1970s. Agile experiments, adaptive programming, and developmental assessment could respond to the challenges of a complex world. Is it enough to reform the development tools or do we need a broader reform of our culture and mindset?

There is a strong view among researchers that all governance thinking should be redefined. Transition management and transformative governance and mission-oriented policymaking are examples of the fairly new systemic governance thinking. These approaches see change being realized through nested arenas and continues patterns of interactions. This means that multilevel governance must simultaneously take into account vertical and horizontal coordination demands by policymakers, stakeholders, and other actors participating in loosely coupled networks. The systems debate in governance settings has started among scholars in the fields of sustainable development and energy policy. It would be desirable if this debate spread to policy areas such as health, education, and foreign and development policies.

Researchers and evaluators play a key role in developing new governance, management, and evaluation practices. Evaluators should act as agile knowledge brokers between decision-makers and citizens. It is therefore important that the functioning of the new management approaches and models is adequately tested before they are introduced. Current transition management models, phenomenon-based policy (cf. Sitra in Finland), and transformative management are still rather descriptive ideal models, and their effectiveness or excellence has not been sufficiently demonstrated by robust empirical research.

In the future, research studies and evaluations based on scientific evidence will be needed. At the same time, however, the term ´scientific evidence´ must be redefined. This must go beyond reductionist scientific ways of thinking, such as overly straightforward causal explanations; instead, it must be equally reliable for interpreting evidence obtained through systems modeling and qualitative research. If, in the future, governance is increasingly based on values and meanings, the applied research and evaluation that support it also need to be able to approach the subject in a new way. This is a challenge for policymakers, researchers, and evaluators alike.

REFERENCES

Abell, P. (2000). Sociological theory and rational choice theory. 223-244. In Turner (ed): The Blackwell Companion to Social Theory 2nd ed. Oxford: Blackwell.

Ahlbæk, E. (1996) 'Why All This Evaluation? Theoretical Notes and Empirical Observation on the Functions and Growth of Evaluation, with Denmark as an Illustrative Case', Canadian Journal of Program Evaluation 11(2), 1–34.

Ahonen, P. (2015). Aspects of the institutionalization of evaluation in Finland: Basic, agency, process and change. Evaluation 21(3), 308–324.

Ahonen, P. and Salminen, A. (1997) Metamorphosis of the Administrative Welfare State. Frankfurt am Main: Peter Lang.

Allen, P. M. (1997). Cities and Regions as Self-organizing Systems. Amsterdam: Gordon & Breach.

Andersson N. A. (2003). Mesoeconomic Analysis of the Construction Sector. Doktorsavhandling, Lund Institute of Technology, Lund University.

Andrews, M., Pritchett, L., Samji, S., and Woolcock, M. (2015). Building Capability by Delivering Results: Putting problem-driven iterative adaptation (PDIA) principles into practice. From A Governance Practitioner's Notebook, retrieved from https://www.oecd.org/dac/accountable- effective-institutions/Governance%20Notebook%202.3%20Andrews%20et%20al.pdf

Ansell, C. & Geyer, R. (2017). 'Pragmatic complexity': A new foundation for moving beyond 'evidence-based policy making'?, Policy Studies, 38(2), 149–167.

Ansell, C., & Gash, A. (2008). Collaborative governance in theory and practice. Journal of Public Administration Research and Theory, 18(4), 543–571.

Aven, T. (2015). Implications of black swans to the foundations and practice of risk assessment and management. Reliability Engineering and System Safety, 134, 83–91.

Bahadur, A.V., Ibrahim, M. & Tanner, T. (2010). The resilience renaissance? Unpacking of resilience for tackling climate change and disasters. Strengthening Climate Resilience. Discussion Paper 1.

Bain, K. A. & Booth, D. & and Wild, L. (2016). Doing Development Differently at the World Bank: Updating the plumbing to fit the architecture. Retrieved from https://www.odi.org/publications/10555-doing-development-differently-world-bank-updating-plumbing-fit-architecture

Bennet, N & Lemoine, G.J. (2014). What VUCA really means for you. Harvard Business Review. Vol. 92. No 1/2. 1-7.

Berkhout, F. & Smith, A. & Stirling, A. (2004). Socio-tecnological regimes and transition contexts, in Elzen, B., Geels, F. and Green, K. (Eds.): System Innovation and the transition to sustainability: Theory, evidence and policy, Edward Elgar, Cheltenham, 48-75.

Bertalanffy, L. (1968): General System Theory – Foundations, Development, Applications. New York: George Braziller.

Bouckaert, G & Halligan, J (2008). Managing Performance. International Comparisons. London: Routledge.

Bouckaert, G. & Van Dooren, W. (2009). Performance Measurement and Management in Public Sector Organizations. In Bovaird, T. & Löffler, E. (Eds.) Public Management and Governance. London: Routledge, 151-164.

Boulton, J. G., Allen, P., Bowman, C. (2015). Embracing Complexity. Strategic Perspectives for an Age of Turbulence. Oxford University Press. Oxford.

Bowen S., Zwi A.B. (2005). Pathways to 'Evidence-Informed' Policy and Practice: A Framework for Action. PLoS Medicine. 2/166. https://doi.org/10.1371/journal.pmed.0020166

Braybrooke, D., & Lindblom, C. E. (1963). A strategy of decision. New York, NY: Free Press.

Brunsson, N. (1987). The Irrational Organization. Irrationality as a Basis for Organizational Action and Change. John Wiley & Sons, Chichester.

Burt, R.S. (2005). Brokerage and Closure: An Introduction to Social Capital. Oxford: Oxford University Press.

Burt, R. S. (1992). Structural Holes. The Social Structure of Competition. Cambridge, MA. Harvard University Press.

Byrne, D. (2013). Evaluating complex social interventions in a complex world. Evsaluation Volume: 19(3), 217-228.

Byrne, D. (2005). Complexity, Configurations and Cases. Theory, Culture and Society 22(5). 95-11.

Byrne, D. (1998). Complexity theory and the social sciences. London: Routledge.

Byrne, D. and Callaghan, G. (2014). Complexity Theory and the Social Sciences: The State of the Art. London: Routledge.

Byrne, D. S. (2011). Applying Social Science. Bristol: Policy Press.

Cairney, P. (2016). The Politics of Evidence-Based Policy Making. London Palgrave.

Caloffi, A. & Rossi, F. & Russo, M. (2015). The emergence of intermediary organizations: a network-based approach to the design of innovation policies. 314- 331. In Geyer, R. & Cairney P.: Handbook on Complexity and Public Policy. Cheltenham UK. EdwardElgar.

Campbell, D.T & Stanley J.C. (1966). Experimental and Quasi-Experimental Designs for Research. Boston: Houghton Mifflin Company.

Campbell D. T. & Russo, M. J. (1999). Social Experimentation. Thousand Oaks: Sage Publications.

Christensen, T., & Laegreid, P. (2016). A Transformative Perspective. 27-43. In Van de Walle, S & Groeneveld, S (Eds.) Theory and Practice of Public Sector Reform. New York: Routledge.

Cilliers, P. (1998). Complexity and Postmodernism. London: Routledge.

Cilliers, P. (2011). Complexity, Poststructuralism and Organisation. 142-154. In Allen, P., Maguire, S., McKelvey, B. (eds.). The Sage Handbook of Complexity and Management. London: Sage Publications.

Cohen, M.D. & March, J.G. (1974). Leadership and Ambiguity: The American College President. New York: McGraw-Hill.

Coleman, J. S. (1990). Foundations in Social Theory. Cambridge, Mass.: The Belknap Press of Harvard University Press.

Cook, Karen S. (1982). Network Structures from an Exchange Perspective. Teoksessa Peter V. Marsden,. Nan Lin (toim.): Social Structure and Network Analysis. Sage, Beverly Hills.

Davies, H., Nutley, S. and Smith, P.C. (2000). What works? Evidence-based policy and practice in public services. London: Policy Press.

Davies, I. C. (1999). Evaluation and performance management in government. Evaluation, 5, 150-159.

Daviter, F. (2017). Coping, taming or solving: Alternative approaches to the governance of wicked problems. Policy Studies, 38(6), 571–588.

Degenne, A. & Forsé, M. (1999). Introducing Social Networks. London. Sage Publications.

Derbyshire, H. & Donovan, E. (2016). Adaptive programming in practice: shared lessons from the DFID-funded LASER and SAVI programmes. Research Paper. August 2016.

Desai, H. & Maneo G. & Pellfolk, E. & Schlingheider, A. (2018). Managing to Adapt. Analysing adaptive management for planning, monitoring, evaluation, and learning. Oxfam Research Reports. March 2018.

Drechsler, W. & Kattel R. (2008). Towards the Neo-Weberian State? Perhaps, but Certainly Adieu, NPM! NISPAcee Journal of Public Administration and Policy 1 (2), 95-99.

Dunsire, A. (1993). Manipulating social tensions: Collibration as an alternative mode of government intervention. MPIfG Discussion Paper 93/7, Max Planck Institute for the Study of Societies.

Durland, M., Fredericks, K. (Eds.), (2005). New Directions for Evaluation: Special Issue: Social Network Analysis in Program Evaluation, Vol. 2005, Issue,107, 1 - 10.

Easley, D. & Kleinberg, J. (2010). Networks, Crowds and Markets. Reasoning about a Highly Connected World. Cambridge: Cambridge University Press.

Elster, J. (1989). Nuts and Bolts for the Social Sciences. Cambridge: Cambridge University Press.

Elzen, B. & Geels, F.W. & Green, K. (Eds.). 2004. System Innovation and the Transition to Sustainability. Theory, Evidence and Policy. Cheltenham: Edward Elgar.

Eoyang G. & Oakden J. (2016). Adaptive Evaluation: A synergy between complexity theory and evaluation practice. Emergence: Complexity and Organization. 2016 Dec 31 [last modified: 2017 Feb 25]. Edition 1. doi: 10.emerg/10.17357.e5389f5715a734817dfbeaf25ab335e5.

Etzioni, A. (1967). Mixed-Scanning: A 'Third' Approach to Decision-Making. Public Administration Review, Vol. 27, No. 5 (Dec., 1967), 385-392.

EVALSED (2008; 2013). The Evaluation of Socio-economic Development. The Guide. Brussels: European Commission.

Fischer, F. (1995). Evaluating Pulic Policies. Chicago: Nelson-Hall Publishers.

Fleischer, N. & Christie, A. (2009). Evaluation Use Results From a Survey of U.S. American Evaluation Association Members. American Journal of Evaluation Vol. 30 Number 2. 2009, 158-175.

Flyvbjerg, B. (2016). What is Phronesis and Phronetic Social Science? Blog in LinkedIn. https://www.linkedin.com/pulse/what-phronesis-phronetic-social-science-bent-flyvbjerg/

Flyvbjerg, B. & Landman, T. & Schram, S. (2012). Important Next Steps in Phronetic Social Science. In: Flyvbjerg, B. & Landman, T. & Schram, S. (Eds.): Real Social Science: Applied Phronesis. Cambridge: Cambridge University Press.

Forss, K. & Schwartz, R. (2011). Introduction in Forss, K. & Marra, M. & Schwartz, R. (Eds.): Evaluating the Complex. Attribution, Contribution, and Beyond. Volume 18. London: Transaction Publishers.

Fox, C. & Grimm, R. & Caldeira, R. (2017). An Introduction to Evaluation. London: Sage.

Freeman, L. C. (1979). Centrality in Social Networks. Conceptual Clarifications, Social Networks (1)/1979, 215-239.

Frej, W. & Ramalingam, B. (2011). Foreign policy and complex adaptive systems: Exploring new paradigms for analysis and action, SFI Working Paper, Sante Fe: SFI http://www.santafe.edu/media/workingpapers/11-06-022.pdf

Fryer, P. (2013). A brief description of Complex Adaptive Systems and Complexity Theory. http://integral-options.blogspot.com/2013/03/peter-fryer-brief-description-of.html

Geels, F.W. (2002). Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case study. Research Policy 31: 8/9, 1257–1274.

Geels, F.W. (2004). Understanding system innovations: a critical literature review and a conceptual synthesis. 19–47. In: Elzen, B., Geels. F. & Green, K. (Eds.). System Innovation and the Transition to Sustainability: Theory, Evidence and Policy. Edward Elgar.

Geyer, R. & Cairney P. (2015). Handbook on Complexity and Public Policy. Cheltenham UK: EdwardElgar.

Gharajedaghi, J. 2011: Systems Thinking. Managing Chaos and Complexity. Amsterdam: Elsevier.

Godenhjelm, S. (2016). Project organisations and Governance - Processes, actors, actions, and participatory procedures. Helsinki: University of Helsinki, Department of Political and Economic Studies.

Goodwin, B. (1995) How the Leopard Changed Its Spots. The Evalution of Complexity. London: Phoenix

Granovetter, M. S. 1973: The Strength of the Weak Ties. American Journal of Sociology. Vol. 78, Issue 6, 1360-1380.

Granovetter, M. S. (1985). Economic Action and Social Structure: The Problem of Embeddedness. American Journal of Sociology 91, 481-510.

Greene, J. G. (1988). Stakeholder participation and utilization in program evaluation. Evaluation Review,12, 91-116.

Guba, E. & Lincoln, Y.S. (1987). The countenances of fourth generation evaluation: Description, judgement and negotiation. In Palumbo, D. J. (Ed.): The Politics of Program Evaluation. Newbury Park, London: Sage.

Hajer, M. A. & Wagenaar, H. (Eds) (2003). Deliberative policy Analysis. understanding Governance in the Network Society. Cambridge: Cambridge University Press.

Hall, P.(1986). Governing the Economy: The Politics of State Intervention in Britain and France. New York: Oxford University Press.

Hanberger , A. (2004). Democratic governance and evaluation. Sixth EES (European Evaluation Society) conference, Berlin, Germany.

Hartley, J. (2005). Innovation in Governance and public services: past and present. Public Money and Management, 25, January, 35–42.

Hayden, G. F. (2006). Policymaking for a Good Society: The Social Fabric Matrix Approach to Policy Analysis and Program Evaluation, New York, NY: Springer.

Hayes, P. (2008). Complexity Theory and Evaluation in Public Management, Public Management Review, 10:3, 401-419.

Head, B. W., & Alford, J. (2015). Wicked problems: Implications for public policy and management. Administration & Society, 47(6), 711–739. Administration & Society, 47(6), 711–739.

Heclo, H. (1978). Issue Networks and the Executive Establishment. In A. King (Ed.): The New American Political System. Washington D.C.: The AEI Press.

Hedström, P. & Swedberg, R. (1996). Rational Choice, Empirical Research, and the

Sociological Tradition. European Sociological Review 12: 127-46.

Hedström, P. & Swedberg, R. (eds.) (1998). Social Mechanisms: An Analytical Approach to Social Theory. Cambridge: Cambridge University Press

Heiskanen, I. and Martikainen, T. (1988). The Finnish Public Sector: Its Growth and Changing Role in 1969-84. In J. A. Lybeck and M. Henrekson (eds). Explaining the Growth of Government. Amsterdam: North-Holland.

Holland, J. (1995) Hidden Order: How Adaptation Builds Complexity. Redwood City, California: Addison Wesley.

Holland, J. (1998) Emergence: From Chaos to Order. Redwood City, California: Addison Wesley.

Holland, J. H. 2014: Complexity. A Very Short Introduction. Oxford: Oxford University Press.

Holling, C.S. (1973). Resilience and Stability of Ecological Systems. Annual Review of Ecology and Systematics, Vol. 4 (1973), pp. 1-23. Available at: http://www.jstor.org/stable/2096802

Hood, C. (1991). A Public Management for All Seasons. Public Administration Vol. 69, 3-19.

Hyytinen, K. (2017). Supporting service innovation via evaluation: a future oriented, systemic and multi-actor approach. Aalto University publication series DOCTORAL DISSERTATIONS, 14/2017, VTT Science, 146

Innes, J. E. & Booher, D. E. (2010). Planning with Complexity. An Introduction to Collaborative Rationality for Public Policy. Routledge. New York.

Jalonen, H. (2007). Kompleksisuusteoreettinen tulkinta hallinnollisen tehokkuuden ja luovuuden yhteensovittamisesta kunnallisen päätöksenteon valmistelutyössä. Tampereen teknillinen yliopisto. Julkaisu 693. Tampere: Yliopistopaino.

Jepperson, R., & Meyer, J. W. (2011). Multiple levels of analysis and the limitations of methodological individualisms. Sociological Theory 29, 54-73.

Johanson, J.-E. & Vakkuri, J. (2017). Governing Hybrid Organisations. Exploring Diversity of Institutional Life. New York: Routledge.

Johanson, J.-E. & Mattila, M. & Uusikylä, P. (1995). Johdatus verkostoanalyysiin. Kuluttujatutkimuskeskus. Menetelmäraportteja ja käsikirjoja. 3/1995, Helsinki.

Kaivo-oja, J. and Stenvall, J. (2013), Foresight, governance and complexity of systems: on the way towards pragmatic governance paradigm, European Integration Studies, Vol. 7 No. 1, 28-34.

Kadushin, C. (2012) Understanding Social Networks. Theories, Concepts and Findings. Oxford University Press. Oxford.

Kaufman, F-X. (1987). Prevention and Intervention in the Analytical Perspective of Guidance. In Klaus Hurrelmann et al. (eds.) Social Interventions: Potential and Constraints. 3-20. Berlin & New York: Walter de Gruyter.

Kaufman, G & Scott, K. E. (2003). What Is Systemic Risk, and Do Bank Regulators Retard or Contribute to It? The Independent Review, v. VII, n. 3, 371– 391.

Kickert, W. and J. Koppenjan (1997). Public Management and Network Management: An Overview. In W. Kickert, E.-H. Klijn and J. Koppenjan (1997) Managing Complex Networks. London: SAGE.

Kirk, R. E. (2013). Experimental design: Procedures for the behavioral sciences (4th ed.). Thousand Oaks, CA: Sage.

Kivimaa, P. & Kern, F. (2016). Creative destruction or mere niche support? Innovation policy mixes for sustainability transitions. Research Policy Vol. 45(1), 2016, 205-217.

Kleinfeld, R. (2015). Improving Development Aid Design and Evaluation: Plan for Sailboats, Not Trains. Washington, DC: Carnegie Endowment for International Peace. http://carnegieendowment.org/files/devt_design_implementation.pdf

Klijn, E-H. & Koppenjan, J. (2012). Governance network theory: past, present and future. Policy & Politics, 40(4), 587–606.

Knight, J. (1994): Institutions and Social Conflict. Cambridge University Press. Cambridge.

Kooiman, J. (2003). Governing as Governance. Sage Publications. London.

Kuhn, T.s S. (1996). The Structure of Scientific Revolution,. The Structure of Scientific Revolutions, 3rd edition. Chicago: University of Chicago Press.

Lakoff, G. (2008). The Political Mind. Why You Can't Understand 21st-Century Politics with an 18th-Century Brain. New York, NY.: Penguin Group.

Lane, J-E. (2000). New Public Management. London, 2000. ProQuest ebrary. Web. 9 March 2015.

Laszlo, E. (1972). Introduction to Systems Philosophy: Toward a New Paradigm of Contemporary Thought. Gordon & Breach Science Publishers.

Leeuw F. (2003). Reconstructing Program Theories: Methods Available and Problems to be Solved. American Journal of Evaluation, n. 24(1), 5-20.

Lehtimäki, H. & Uusikylä, P. & Smedlund, A. (Eds.) (2020). Society as an Interaction Space. A Systemic Approach. Translational Systems Sciences. Singapore: Springer. Forthcoming.

Lindblom, C. (1959). The science of muddling through. Public Administration Review, 19(2), 79–88.

Loorbach, D. (2009). Transition Management for Sustainable Development: A Prescriptive, Complexity-Based Governance Framework. Governance: An International Journal of Policy, Administration, and Institutions, Vol. 23(1) , 161–183.

Loorbach, D. & Rotmans, J. (2006). Managing Transitions for Sustainable Development. In: Olsthoorn, Xander, Wieczorek, Anna J. (Eds.): Understanding Industrial Transformation - Views from different disciplines, 187-206. Amsterdam: Springer.

Luhman, N. (1995). Social Systems. California: Stanford University Press.

Lynn, L. (2008). What is a Neo-Weberian State? Reflections on a Concept and its Implications. NISPAcee Journal of Public Administration and Policy 1 (2), 17-30.

Lähteenmäki-Smith, K. & Virtanen, P. (2018). Societal Phenomena, Mission-oriented Public Policy and the New Evaluation Culture. Paper presented at the EES Conference in Thessaloniki. October 2-5, 2018.

MacDonald,B. (1979). Democracy and Evaluation. Centre for Applied Research in Education. University of East Anglia. https://www.uea.ac.uk/documents/4059364/4410085/MacDonald-1979-Democracy+and+Evaluation.pdf/b9cabf45-aef1-439e-91de-aba10aa34760

March, J.G. and Olsen, J.P. (1976). Ambiguity and Choice in Organizations. American Journal of Sociology, 84, 765-767

Mattila, M. (2000). Policy Making in Finnish Social and Health Care – A Network Approach. Acta Politica 12. Department of Political Science, University of Helsinki.

Mayne, J. (2011). Addressing Cause and Effect in Simple and Complex Settings through Contribution Analysis. In Evaluating the Complex, R. Schwartz, K. Forss, and M. Marra (Eds.). New York: Transaction Publishers.

Mayntz, R. (1993). Governing Failures and the Problem of Governability: Some Comments on a Theoretical Paradigm. In J. Kooiman (ed.): Modern Governance. New Government-Society Interactions. London: Sage Publications.

Mazzucato, M. (2018a). Mission-Oriented Research & Innovation in the European Union. A problem-solving approach to fuel innovation-led growth. European Commission. Directorate-General for Research and Innovation. Brussels.

Mazzucato, M. (2018b). The Value of Everything. Making and Taking in the Global Economy. London: Penguin Random House.

Mazzucato, M. (2013), The Entrepreneurial State: Debunking the Public Vs. Private Myth in Risk and Innovation. London: Anthem Press.

McGuire, M. (2006) Collaborative Public Management: Assessing What We Know and How We Know It. Public Administration Review. Public Administration Review. Vol. 66, Special Issue: Collaborative Public Management (Dec., 2006), 33-43.

Meadowcroft, J. (2011). Engaging with the politics of sustainability transitions, Environmental Innovations and Societal Transitions, 1, 70-75.

Metcalfe, L. & Richards, S. (1990). lmproving Public Management. London: Sage.

Meyer, M. (2010). The Rise of the Knowledge Broker. Science Communication, Vol. 32, (1), 118-127.

Mohr, Lawrence B. (1995). Impact Analysis for Program Evaluation. Thousand Oaks: Sage.

Miedzinski, M. & Mazzucato, M. & Ekins, M. (2019). A framework for mission-oriented innovation policy roadmapping for the SDGs: The case of plastic-free oceans. Director, UCL Institute for Innovation and Public Purpose. Working Paper 2019-3.

Monge, P. R., & Contractor, N. S. (2008). Emergence of communication networks. In L. Putnam & F. Jablin (Eds.) New handbook of organizational communication. Newbury Park, CA: Sage

Montague S.(2009). Structured Contribution Analysis A Brief Dialogue and Practical Demonstration. Presentation at the University of Edinburg.

Newman, J. (2017). Deconstructing the debate over evidence-based policy. Critical Policy Studies 11(2): 211–226.

Newman, J. & Head, B.W. (2017) Wicked tendencies in policy problems: rethinking the distinction between social and technical problems. Policy and Society, 36(3): 414–429.

Nutley, S. & Walter, I & Davies, H. (2003). From knowing to doing: a framework for understanding the evidence-into-practice agenda. Evaluation 9(2). 125-48.

Nutley, S. (2003). Bridging the policy/research divide: Reflections and Lessons from the UK, Keynote paper at National Institute of Governance Conference; Facing the Future: Engaging stakeholders and citizens in developing public policy', Canberra, Australia 23/24 April 2003.

OECD (1998). Best Practice Guidelines for Evaluation. PUMA Policy Brief No. 5. Paris: OECD.

Ouchi, W. G., (1980) "Markets, Bureaucracies, and Clans," Administrative Science Quarterly, 25 (March), 129-141.

Oxfam 2013: Toward Resilience. A Guide to Risk Reduction and Climate Change Adaptation. Practical Action Publishing. Warwickshire.

Parsons, T (1977). On Building Social Systems Theory. In Parsons, Talcott: Social Systems and the Evolution of Action Theory. Free Press. New York.

Patton M. Q. (2015). The sociological roots of utilization-focused evaluation. The American Sociologist 4(46): 457 – 462.

Patton, M. Q. 2011: Developmental Evaluations. Applying Complexity Concepts to Enhance Innovation and Use. New York. Guilford.

Patton, M. Q. (1997) Utilization-focused Evaluation. Newbury Park, CA: Sage.

Patton, M. Q. (1998) 'Discovering Process Use', Evaluation 4(2): 225–33.

Patton, M.Q., McKegg, K. & Wehipeihana, N. (eds.) (2016). Developmental Evaluation Exemplars: Principles in Practice. New York. Guilford.

Pawson, R. (2006). Evidence-based Policy. A Realistic Perspective. London: SAGE Publications.

Pawson, R. (2013). The Science of Evaluation. A Realist Manifesto. Sage Publications. Los Angeles.

Pawson, R. & Tilley, N. (1997). Realistic evaluation. Sage, Thousand Oaks.

Pierre, J. & Peters G. P. (2000). Governance, Politics and the State. MacMillan Press. Houndmills.

Polanyi, M. (1967). The Tacit Dimension. Doubleday-Anchor. GardenCity.

Pollit, C. and G. Bouckaert (2003) Evaluating Publie Management Reforms: An International Perspective', in H. Wollmann (ed.) Evaluation in Publie-Seetor Reform. Coneepts and Praetiee in International Perspeetive, pp. 12-35. Cheltenham: Edward Elgar.

Pollitt, C. (1990). Managerialism and the public services The Anglo-American experience. Oxford Basil Blackwell.

Polllit, C. & Bouckaert, G. (2011). Public Management Reform. A Comparative Analysis: NPM, Governance and the Neo-Weberian State. Oxford: Oxford University Press.

Preskill, H & Beer T. (2012). Evaluating Social Innovation. Center for Evaluation Innovation

Provan, K. G. & Kenis, P. (2008). Modes of Network Governance: Structure, Management, and Effectiveness". Journal of Public Administration.

Puustinen, A. (2017). Voiko verkostoa johtaa? Tapaustutkimus sosiaali- ja terveydenhuollon yhteistoiminta-alueen hallinnan yhteenkietoutuneesta luonteesta. [Can networks be managed? Case study on the complexity of governance in a co- operation district of health and social care.] Publications of the University of Eastern Finland Dissertations in Social Sciences and Business Studies; 149.

Radej, B. (2015). Social Complexity: Operational definition. Slovenian Evaluation Society - Working Papers 7/2 (July 2015).

Raisio, H. & Jalonen, H. & Uusikylä, P. (2018). Kesy, sotkuinen vai pirullinen ongelma. Tiedon käyttö yhteiskunnallisessa päätöksenteossa. Sitran selvityksiä 139. Marraskuu 2018. Sitra. Helsinki.

Raisio, H., Puustinen, A. & Vartiainen, P. (2018b). The Concept of Wicked Problems: Improving the Understanding of Managing Problem Wickedness in Health and Social Care. Teoksessa Thomas, W., Hujala, A., Laulainen, S. & McMurray, R. (toim.) The Management of Wicked Problems in Health and Social Care. Lontoo: Routledge.

Rajavaara M. (2007). Vaikuttavuusyhteiskunta. Sosiaalisten olojen arvostelusta vaikutusten todentamiseen. Helsinki: Kela, Tutkimusosasto, Sosiaali- ja terveysturvan tutkimuksia 84, 2007.

Ramalingam, B. 2013: Aid on the Edge of Chaos. Oxford University Press. Oxford.

Reynolds, M. (2015). (Breaking) The Iron Triangle of evaluation. IDS Bulletin 46(1), 71-86.

Richardson, K. (2008). Managing complex organizations: Complexity thinking and the science and art of management. E:CO, 10(2): 13–26.

Richardson, K. & Cilliers, P. (2001). Special editors' introduction: What is complexity science? A view from different direction. Emergence, 3(1): 5–23.

Richmond, B. 1994: System Dynamics/Systems Thinking: Let's Just Get On With It". A paper presented at the 1994 International Systems Dynamics Conference, Sterling, Scotland.

Rittel, H.W.J., & Webber, M. (1973). Dilemmas in a general theory of planning. Policy Sciences, 4(2): 155–169.

Rotmans, J. & Kemp, R & van Asselt, M. (2001). More evolution than revolution: transition management in public policy. Foresight, Vol. 3 Issue: 1, pp.15-31.

Sabatier, P. (1993): Policy Change over a Decade or More. In Sabatier, P. and Jenkins-Smith H. (Eds.) Policy Change and Learning. An Advocacy Coalition Approach. Boulder: Westview Press.

Schot, J. & Steinmueller, W. E. (2018). Three Frames for Innovation Policy. R&D, Systems of Innovation and Transformative Change. Research Policy 47 (2018). 1554-1567.

Scott, J. (1991): Social Network Analysis. Introduction. London: Sage.

Scott, W. R. (1995). Institutions and Organizations. Sage Publications. Thousand Oaks.

Segone, M. (2004). Bridging the gap, the role of monitoring and evaluation in Evidence-based Policymaking . UNICEF.

Senge, P. (2006). The Fifth Discipline. The Art & Practice of the Learning Organization. Random House. New York.

Setälä, M. (2017): Connecting Deliberative Mini-Publics to Representative Decision- Making. European Journal of Political Research 56, 846–863.

Simon, H. A. (1957): Administrative Behavior. A Study of Decision Making Process in Administrative Organization. Second Edition. Free Press, New York.

Sitra (2018). Ilmiömäinen julkinen hallinto. Keskustelualoite valtioneuvoston toimintatapojen uudistamiseksi. Sitran työpaperi 31.8.2018.

Sjöblom, S., Löfgren, K., & Godenhjelm, S. (2013). Projectified Politics – Temporary Organisations in Public Context Introduction to the special issue. Scandinavian Journal of Public Administration, 17(2), 3–11.

Skocpol, T. (1985). Bringing the State back in. The strategies of analysis of Current Research. Edited with Peter Evans. Cambridge University Press.

Smith, A. & Stirling, A. & Berkhout, F. (2005). The governance of sustainable socio- technical transitions, Research Policy, 34, 1491-1510.

Snowden, D.J. & Boone M.E. (2007) A Leader´s Framework for Decision Making. Harvard Business Review 85(11). 68-77.

Stacey, R. D. (1996). Complexity and Creativity in Organizations. San Francisco: Berrett-Koehler Publishers.

Stacey, R.D. & Mowles, C. (2016). Strategic Management and Organisational Dynamics. The challenge of complexity to ways of thinking about organisations. Seventh Edition. New York: Pearson.

Stame, N. (2004). Theory-Based Evaluation and Types of Complexity. Evaluation 10(1), 58-76.

Stroh, D. P. (2015). Systems Thinking for Social Change. A Practical Guide to Solving Complex Problems, Avoiding Unintended Consequences, and Achieving Lasting Results. Chelsea Green Publishing. White River Junction, Vermont.

Ståhle P. (2004) Itseuudistumisen dynamiikka – Systeemiajattelu kehitysprosessien ymmärtämisen perustana. Teoksessa: Sotarauta M. & Kosonen K.-J.(toim.) Yksilö, kulttuuri, innovaatioympäristö. S. 222–255. University Press.Tampere: Tampere

Sutcliffe, S. & Court, J. (2005). Evidence-Based Policymaking: What is it? How does it work? What relevance for developing countries? Overseas Development Institute. November 2005.

Sweeney, L. B., & Sterman, J. D. (2000). Bathtub dynamics: initial results of a systems thinking inventory. System Dynamics Review, 16(4), 249–286.

Temmes, M. (1998). Finland and New Public Management. International Review of Administrative Sciences. Vol. 64 (3). 441-456.

Termeer, C. J. A. M. & Dewulf, A. (2018). A small wins framework to overcome the evaluation paradox of governing wicked problems. Policy and Society, online first.

Thompson, D.F. (2008). Deliberative Democratic Theory and Empirical Political Science. Annual Review of Political Science 11: 497-520.

Thomson, A. & Perry, J. (2006) Collaboration Processes: Inside the Black Box. Public Administration Review. 66. 20 - 32.

Tiihonen, S. (2004). From Governing to Governance. A Process of Change. Tampere University Pres. Tampere.

Torfing, J., Peters, B. G., Pierre, J., & Sorensen, E. (2013). Interactive Governance: Advancing the Paradigm. Oxford University Press.

Tuomi, I. (2018). The Impact of Artificial Intelligence on Learning, Teaching, and Education. Policies for the future, Eds. Cabrera, M., Vuorikari, R & Punie, Y., EUR 29442 EN, Publications Office of the European Union, Luxembourg.

Uusikylä, P. (2013). Transforming Silo-steering into Performance Governance System: The Case of The Finnish National Government. New Directions for Evaluation, 137.

Uusikylä, P., Tommila, P. & Uusikylä, I. (2020). Disaster Management as a Complex Adaptive System. In Lehtimäki, H. & Uusikylä, P. & Smedlund, A. (Eds.). Society as an Interaction Space. A Systemic Approach. Translational Systems Sciences. Singapore: Springer. Forthcoming.

Uusikylä, P. & Pulkkinen, J. & Pesonen, P. & Salminen, V. (2016). Complexity Calls for a New Evaluation Paradigm. Lessons learned from the BEAM Program. Paper proposal for the 12th EES Biennial Conference (EES 2016) "Evaluation Futures in Europe and beyond: Connectivity, Innovation and Use" September 28-30, 2016, Maastricht, The Netherlands

Uusikylä, P. & Tommila, P. & Rinne, P. & Valtari, H. (2017). FRC International Disaster Preparedness and Disaster Risk Reduction (DPDRR) Programmes Thematic Study. Finnish Red Cross. Helsinki.

Uusikylä, P. & Virtanen, P. (2008). Evaluators as Information Brokers – Approach to enhance social capital and policy coherence. Conference paper in EES Conference, 1-3 October 2008. STREAM 4 Methodological Choices and Approaches.

Uusikylä, P. & Virtanen, P. (2000). Metaevaluation as a Tool for Learning: A Case Study of the European Structural Fund Evaluations in Finland. Evaluation 1 (6), 50-65.

Uusikylä, P. & Valovirta, V. (2007). Three Spheres of Performance Governance: Spanning the boundaries from single-organisation focus towards a partnership network', Evaluation 13(4).

Valovirta, V. (2002). Evaluation Utilization as Argumentation. January 2002. Evaluation 8(1):60-80.

Vedung, E. (1997). Public Policy and Program Evaluation. Transaction Publiahers, New Jersey.

Vento, I. (2017). The Evaluation of Innovations : How to Evaluate the Transformative Value of the Cohesion Policy for Managing Purposes. Scandinavian Journal of Public Administration, 21(2), 3–22.

Virtanen, Petri (2007). Arviointi. Arviointitiedon luonne, tuottaminen ja hyödyntäminen. Helsinki: Edita.

Virtanen, P. & Kaivo-oja, J. (2015). Public service systems and emerging systemic governance challenges", International Journal of Public Leadership, Vol. 11 Iss 2 pp. 77 - 91

Virtanen, P. & Stenvall, J. (2018). Intelligent Health Policy. Theory, concept and practice. Cham: Springer.

Virtanen, P. & Uusikylä, P. & Jalava, J & Tiihonen, S. & Laitinen, L. & Noro, K. (2016). Valtioneuvoston yhtenäisyys – kansainvälinen vertaileva tutkimus. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 49/2016.

Virtanen, P. & Uusikylä, P. (2004). Exploring the Missing Links Between Causes and Effects. A Conceptual Framework for Understanding Micro-Macro Conversions in Programme Evaluation. 10(1). Evaluation. SAGE Publications.

Ward, V. & House, A. & Hamer, S. (2009). Knowledge brokering: the missing link in the evidence to action chain? Evidence & Policy: A Journal of Research, Debate and Practice, Volume 5, 13(3), pp. 267-279.

Wahl, D.C. (2017). The adaptive cycle as a dynamic map for resilience thinking. Medium. Available at: https://medium.com/@designforsustainability/the-adaptive-cycle-panarchy-as-dynamic-maps-for-resilience-thinking-793fad49de5e

Wallenius, T. & Uusikylä, P. & Tran Mihn, A. (2019). Final Evaluation of the Innovation Partnership Programme, Phase II. Ministry for Foreign Affairs, Finland. ASA-10. May 2019.

Walton, M. (2016). Setting the context for using complexity theory in evaluation: boundaries, governance and utilisation. Evidence & Policy (12) 1. 73–89.

Wasserman, S. and Faust, K. (1994). Social Network Analysis. Cambridge: Cambridge University Press.

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. Nature. 393 (6684): 440–442.

Weber, K. M. & Rohracher, H. (2012). Legitimizing research, technology and innovation policies for transformative change. Combining insights from innovation systems and multi-level perspective in a comprehensive ´failures´framework. Research Policy 41 (2012). 1037-1047.

WEF (2014). Global Risks 2014. Ninth Edition. World Economic Forum. Geneva.

Weick, K. & Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. Organization Science, 16(4): 409–421.

Weiss, C. H. (1979). The many meanings of research utilization. Public Administration Review 39, 426-431.

Weiss, C. H. (1998a). Evaluation. New Jersey: Prentice Hall.

Weiss, C. H. (1998b). Evaluation: Methods for Studying Programs and Policies. Prentice Hall. Upper Saddle River, NJ.

Weiss, Carol H. (1997). Theory-Based Evaluation Past, Present, and Future. New Directions for Evaluation 76, 41-55.

White H. (2009). Theory-Based Impact Evaluation: Principles And Practice, Working Paper n. 3, International Initiative for Impact Evaluation, New Delhi.

White, H., & S. Sabarwal (2014). Quasi-experimental Design and Methods, Methodological Briefs: Impact Evaluation 8, UNICEF Office of Research, Florence.

Wiener, N. (1950). The Human Use of Human Beings. The Riverside: Houghton Mifflin Co.

Williams, B., & Hummelbrunner, R. (2011). Systems concepts in action: A practitioner's toolkit. Stanford, CA: Stanford University Press.

Williams, Bob & Hummelbrunner, Richard (2009). Outcome Mapping 75-91. In Williams, Bob & Hummelbrunner, Richard. Systems Concepts in Action. A Practitioner´s Toolkit. Stanford University Press. Stanforg, CA.

Williamson, O. E., (1979): Transaction-Cost Economics: The Governance of Contractual Relations, Journal of Law and Economics, 22(2), 233-261.

Wilson-Grau, R. (2015). Outcome Harvesting. BetterEvaluation. Retrieved from http://betterevaluation.org/plan/approach/outcome_harvesting

Wimbush, E. & Montague S. & Mulherin, T. (2012). Applications of contribution analysis to outcome planning and impact evaluation. Evaluation 3/2012. 310-329.

World Bank (2001). Understanding And Measuring Social Capital: A Synthesis Of Findings And Recommendations From The Social Capital Initiative. Social Capital Initiative Working Paper No. 24. Washington D.C.: The World Bank.

Wrong, D. (1961): The Oversocialized Conception of Man in Modern Sociology. American Sociological Review 26 (2): 183-93.

Ylikoski, P. (2016). Thinking with the Coleman Boat. The Institute for Analytical Sociology. The IAS Working Paper Series 2016: 1. Linköping: Linköping University.